

UNIVERSITÄT MANNHEIM  
FAKULTÄT FÜR SOZIALWISSENSCHAFTEN  
FACHBEREICH PSYCHOLOGIE  
LEHRSTUHL PÄDAGOGISCHE PSYCHOLOGIE



Unterrichtsqualität aus der Schülerperspektive:  
Ein Verfahren zur Unterrichtsentwicklung

Diplomarbeit

vorgelegt von  
stud. cand. psych.

Kathleen Piskol

Römerstr. 28a  
69226 Nußloch

9. Oktober 2008

Erstgutachter: Prof. Dr. Manfred Hofer

Zweitgutachterin und Betreuerin: Dr. Birgit Pikowsky

## Inhaltsverzeichnis

<b>Abbildungsverzeichnis .....</b>	<b>V</b>
<b>Tabellenverzeichnis .....</b>	<b>VIII</b>
<b>Zusammenfassung .....</b>	<b>1</b>
<b>1 Einleitung .....</b>	<b>2</b>
1.1 Fragestellung und institutionelle Einbettung .....	3
1.2 Zielsetzung und Aufbau .....	5
<b>I Theoretischer Teil .....</b>	<b>7</b>
<b>2 Unterrichtsqualität als Meta-Konstrukt .....</b>	<b>7</b>
2.1 Konzeptualisierung von Unterrichtsqualität .....	7
2.2 Modelle und Befunde der Unterrichtsforschung .....	12
2.2.1 Carrolls und Blooms Modell schulischen Lernens .....	12
2.2.2 Weiterentwicklungen durch Creemers und Slavin .....	14
2.2.3 Walbergs Produktivitätsmodell .....	16
2.2.4 Angebot-Nutzungs-Modell unterrichtlicher Wirkungen von Helmke .....	17
2.3 Strukturierungsansätze zu den Dimensionen der Unterrichtsqualität .....	20
<b>3 Unterrichtswahrnehmung aus Schülersicht .....</b>	<b>24</b>
3.1 Schülerwahrnehmung als Fokus der Klimaforschung .....	24
3.1.1 Wurzeln der Klimaforschung .....	24
3.1.2 Schülerwahrnehmung als eine Konstituente des Klimabegriffs .....	26
3.2 Schülerwahrnehmung als interpersonaler Wahrnehmungsprozess .....	28
3.3 Schüler als Datenquelle .....	30
3.4 Operationalisierung von Schülerwahrnehmungen .....	34
3.4.1 Itemformulierung .....	35
3.4.2 Antwortformat .....	38
3.4.3 Urteilsbezug .....	39
<b>4 Schülerfeedback als Beitrag zur Unterrichtsentwicklung .....</b>	<b>40</b>
4.1 Evaluation als Funktion von Schülerfeedback .....	40
4.2 Determinanten der Wirksamkeit von Selbstevaluation .....	44
4.3 Implementierung des Schülerfeedbackverfahrens .....	47
4.3.1 Ebene der Unterrichtsentwicklung .....	47
4.3.2 Ebene der Personal- und Organisationsentwicklung .....	52
<b>5 Präzisierung der Zielsetzungen .....</b>	<b>54</b>
<b>II Empirischer Teil .....</b>	<b>57</b>
<b>6 Qualitative Voruntersuchung .....</b>	<b>57</b>
6.1 Durchführung kognitiver Pretestinterviews .....	57
6.2 Darstellung der Interviewergebnisse .....	60

<b>7</b>	<b>Haupterhebung .....</b>	<b>63</b>
7.1	Durchführung der Befragung.....	63
7.2	Aufbereitung der Rohdaten.....	65
7.3	Beschreibung der Stichproben .....	67
7.4	Ergebnisrückmeldung an die Lehrkräfte.....	69
<b>8</b>	<b>Empirische Vorarbeiten .....</b>	<b>71</b>
8.1	Auswertungsvorgehen .....	71
8.2	Prüfung der Analysevoraussetzungen.....	72
<b>9</b>	<b>Analyse der Items zur Unterrichtswahrnehmung .....</b>	<b>77</b>
9.1	Deskriptivstatistische Befunde .....	77
9.1.1	Auswertungsvorgehen.....	77
9.1.2	Itemmissinganalyse .....	78
9.1.3	Verteilungsanalyse .....	79
9.1.4	Analyse der Lokationsmaße .....	82
9.1.5	Analyse der Dispersionsmaße .....	88
9.1.6	Zusammenfassende Bewertung .....	91
9.2	Item- und Reliabilitätsanalysen .....	91
9.2.1	Auswertungsvorgehen.....	92
9.2.2	Skala Klassenführung.....	94
9.2.3	Skala Lernförderliches Klima .....	96
9.2.4	Skala Motivierung .....	97
9.2.5	Skala Klarheit und Verständlichkeit.....	98
9.2.6	Skala Aktivierung.....	99
9.2.7	Zusammenfassende Bewertung .....	100
9.3	Faktorielle Validitätsanalysen.....	102
9.3.1	Auswertungsvorgehen.....	102
9.3.2	Exploratorische Faktorenanalyse .....	103
9.3.3	Konfirmatorische Faktorenanalyse.....	106
9.3.4	Zusammenfassende Bewertung .....	111
9.4	Weitere Validitätsanalysen .....	112
9.4.1	Vorgehen .....	112
9.4.2	Inhaltsvalidität .....	113
9.4.3	Augenscheinvalidität .....	113
9.4.4	Exkurs: Übereinstimmung zwischen Schüler- und Lehrerangaben .....	115
9.4.5	Zusammenfassende Bewertung .....	116
<b>10</b>	<b>Analyse der Angaben zur Urteilssicherheit .....</b>	<b>117</b>
<b>11</b>	<b>Itemselektion .....</b>	<b>120</b>
11.1	Eignung der ich- vs. wir-Version.....	120
11.2	Eignung auf Itemebene .....	121

---

<b>12 Diskussion .....</b>	<b>126</b>
12.1 Methodische Überlegungen .....	126
12.2 Kritische Bewertung des Fragebogens .....	130
12.3 Weiterführende Forschungs- und Anwendungsfragen .....	134
12.4 Resümee und Ausblick .....	137
<b>Literaturverzeichnis.....</b>	<b>139</b>
<b>Anhang .....</b>	<b>I</b>

### Abbildungsverzeichnis

Abbildung 1:	<i>Darstellung der institutionellen Einbettung dieser Arbeit</i>	5
Abbildung 2:	<i>Carrolls Modell schulischen Lernens (1964) in Anlehnung an Harnischfeger und Wiley (1977, S. 209)</i>	12
Abbildung 3:	<i>Slavins QAIT-Modell (in Anlehnung an Slavin, 1994, S. 149)</i>	15
Abbildung 4:	<i>Produktivitätsmodell schulischen Lernens nach Walberg (in Anlehnung an Fraser, Walberg, Welch &amp; Hattie, 1987, S. 158)</i>	17
Abbildung 5:	<i>Angebots-Nutzungs-Modell von Helmke (2007c) mit Hervorhebung des im Fragebogen erfassten Modellausschnitts</i>	18
Abbildung 6:	<i>Erweitertes Modell sozialisatorischer Einflüsse (Getzels &amp; Thelen, 1960)</i>	25
Abbildung 7:	<i>Schülerfeedbackschleife zum Einsatz des Fragebogentandems für Lehrkräfte im Rahmen der Unterrichtsentwicklung</i>	48
Abbildung 8:	<i>Feedbackrad sich gegenseitig ergänzender Feedbackquellen für das komplexe Unterrichtsgeschehen, wobei das Fragebogentandem des Projekts die Selbst- und Fremdeinschätzung durch die Schüler abdeckt</i>	54
Abbildung 9:	<i>Stichprobenzusammensetzung innerhalb der verschiedenen Erhebungsphasen</i>	66
Abbildung 10:	<i>Relation der Missinganzahl des Items SKF10 zu der von Item SKV55 mit einer durchschnittlichen Anzahl von 17 Missings über alle erfassten Klassen hinweg</i>	78
Abbildung 11:	<i>Histogramme zum linksschiefen Item SKF1, rechtsschiefen Item SA72, zum eine verzerrte Normalverteilungsform aufweisenden Item SM37, zum gleichverteilten Item SA67 und zum zweigipflig verteilten Item SKF12 sowie Angabe der Häufigkeiten der jeweiligen Verteilungsformen in der Gesamtstichprobe (<math>H_{Items}</math>)</i>	81
Abbildung 12:	<i>Verteilung der Itemmittelwerte zur Unterrichtswahrnehmung innerhalb der theoretisch postulierten Unterrichtsmerkmale (KF = Klassenführung, LK = Lernförderliches Klima, M = Motivierung, KV = Klarheit und Verständlichkeit, A = Aktivierung) mit Angabe der 95%-Konfidenzintervalle</i>	84
Abbildung 13:	<i>Verteilung der Itemmittelwerte in Abhängigkeit von der Fragebogenversion sowie des Unterrichtsmerkmals (KF = Klassenführung, LK = Lernförderliches Klima, M = Motivierung, KV = Klarheit und Verständlichkeit, A = Aktivierung)</i>	85
Abbildung 14:	<i>Darstellung der Itemstandardabweichungen zur Unterrichtswahrnehmung in Abhängigkeit von der Fragebogenversion sowie des Unterrichtsmerkmals (KF = Klassenführung, LK = Lernförderliches Klima, M = Motivierung, KV = Klarheit und Verständlichkeit, A = Aktivierung)</i>	89

Abbildung 15:	<i>Dargestellt sind die gegen den mehrstufigen Zustimmungssindex geplotteten Trennschärfen der Klassenführungsitems in Abhängigkeit von der Fragebogenversion</i>	95
Abbildung 16:	<i>Dargestellt sind die gegen den mehrstufigen Zustimmungssindex geplotteten Trennschärfen der Items zur Erfassung des Merkmals Lernförderliches Klima in Abhängigkeit von der Fragebogenversion</i>	96
Abbildung 17:	<i>Dargestellt sind die gegen den mehrstufigen Zustimmungssindex geplotteten Trennschärfen der Motivierungsitems in Abhängigkeit von der Fragebogenversion</i>	97
Abbildung 18:	<i>Dargestellt sind die separat für die Fragebogenversionen sowie für die Gesamtstichprobe geplotteten mehrstufigen Zustimmungssindizes und die Trennschärfen der Items zur Erfassung der Klarheit und Verständlichkeit</i>	98
Abbildung 19:	<i>Plot der mehrstufigen Zustimmungssindizes der Aktivierungsitems gegen die Trennschärfewerte für die beiden Fragebogenversionen sowie in der Gesamtstichprobe</i>	100
Abbildung 20:	<i>Scree-Plot der 78 Faktoren gegen die Eigenwerte, wobei die Hilfslinie eine Gerade darstellt, die durch die letzten extrahierten Faktoren gelegt wird, die von Cattell (1966) als ‚Geröll‘ bezeichnet werden</i>	103
Abbildung 21:	<i>Vergleich der empirischen und des aufgrund der Parallelanalyse nach Horn (1965) ermittelten Eigenwertverlaufs der Zufallszahlen, wobei der Pfeil den 10. Faktor markiert, bei dem der empirische Eigenwert zum ersten Mal kleiner ist als der auf der Grundlage von Zufallszahlen ermittelte Eigenwert</i>	104
Abbildung 22:	<i>Dargestellt ist das a priori postulierte 5-Faktoren-Mess- bzw. Strukturmodell der 78 Items zur Unterrichtswahrnehmung mit den entsprechenden Fit-Indizes</i>	108
Abbildung 23:	<i>Gegenüberstellung der Itemmittelwerte für die Lehrer- vs. Schülerstichprobe</i>	116
Abbildung 24:	<i>Verteilung der Itemmittelwerte der Urteilssicherheit innerhalb der theoretisch postulierten Unterrichtsmerkmalsbereiche (KF = Klassenführung, LK = Lernförderliches Klima, M = Motivierung, KV = Klarheit und Verständlichkeit, A = Aktivierung)</i>	117
Abbildung 25:	<i>Verteilung der Itemmittelwerte in Abhängigkeit von der ich- vs. wir-Version</i>	118
Abbildung 26:	<i>Standardabweichungen der Urteilssicherheitsangaben in Abhängigkeit vom Item und der Fragebogenversion</i>	119
Abbildung C.1-1:	<i>Rahmenplan zur Stichprobenakquise</i>	XIV
Abbildung D.3-2:	<i>Konfirmatorische 5-Faktorenlösung ohne Items mit exploratorischer Faktorladung kleiner .30</i>	LI
Abbildung D.3-3:	<i>Mess- bzw. Strukturmodell der Schmid-Leiman-Lösung</i>	LII

## Tabellenverzeichnis

Tabelle 1:	<i>Gegenüberstellung der von Clausen et al. (2002) ermittelten Clustergruppen sowie beispielhafter Indikatoren zu den von Helmke (2006b) postulierten 10 Merkmalen guten Unterrichts, wobei die im Fragebogen erfassten Merkmale hervorgehoben sind</i>	22
Tabelle 2:	<i>Facetten zur Eingrenzung des Klimabegriffs (Eder, 1996; 2001)</i>	26
Tabelle 3:	<i>4-Felder-Matrix zur Wahrnehmungsperspektive vs. Adressat des Lehrerverhaltens sowie der Zuordnung von Eders Klassifikation (2002, S. 219) anhand des Items 21 im Projektfragebogen (in Anlehnung an Wagner, 2008, S. 9)</i>	36
Tabelle 4:	<i>Hauptfunktionen der Evaluation (in Anlehnung an Hense, 2006, S. 51)</i>	43
Tabelle 5:	<i>Überblick zu den Zielsetzungen der Arbeit</i>	56
Tabelle 6:	<i>Verteilung der Stichprobenelemente der qualitativen Voruntersuchung hinsichtlich der besuchten Klassenstufe, Schulart und der Deutschkenntnisse</i>	59
Tabelle 7:	<i>Übersicht zu Items mit Validitätsproblemen in der SFB-Pretestversion sowie deren Umformulierung bzw. Elimination hinsichtlich der SFB-Überprüfungsversion</i>	62
Tabelle 8:	<i>Verteilung der Lehrkräfte auf die Schulart und die Klassenstufe.</i>	67
Tabelle 9:	<i>Anzahl der eingeschätzten Unterrichtsfächer in Abhängigkeit von den verschiedenen Klassenstufen</i>	68
Tabelle 10:	<i>Verteilung der Produktmomentkorrelationen innerhalb der gesamten Schülerstichprobe</i>	76
Tabelle 11:	<i>Spannweiten negativer und positiver Schiefewerte für die Fragebogenversionen sowie insgesamt</i>	79
Tabelle 12:	<i>Spannweiten negativer und positiver Kurtosiswerte für die Fragebogenversionen sowie insgesamt</i>	80
Tabelle 13:	<i>Relative mittlere Häufigkeit pro Antwortkategorie in der Gesamtstichprobe sowie den Teilstichproben der ich- und wir-Version</i>	88
Tabelle 14:	<i>Reliabilitäts- und Homogenitätswerte der fünf Skalen in Abhängigkeit von der Fragebogenversion</i>	101
Tabelle 15:	<i>Items mit standardisierten Regressionsgewichten <math>\leq .35</math></i>	109
Tabelle 16:	<i>Modell-Fit-Indizes zu verschiedenen Modellvarianten</i>	110
Tabelle 17:	<i>Interkorrelationsmatrix der Unterrichtsmerkmalsskalen</i>	110
Tabelle 18:	<i>Modell-Fit-Indizes zu verschiedenen Modellvarianten auf der Gesamtebene</i>	111
Tabelle 19:	<i>Qualitative Angaben der Haupterhebung, die hinsichtlich der Augenscheinvalidität auffällig sind</i>	114
Tabelle 20:	<i>Ergebnisgegenüberstellung für die ich- vs. wir-Version</i>	121

Tabelle 21:	<i>Zusammenstellung von auffälligen Items mit den relevanten Gütekriterien</i>	123
Tabelle A.1-1	<i>Kurzbeschreibung zum Fragebogen</i>	I
Tabelle C.2-2	<i>Kreuztabellierte Häufigkeitsverteilung der Missings für die Items zur Unterrichtswahrnehmung (NMISS_UNT) und Urteilssicherheit (NMISS_US)</i>	XXVI
Tabelle C.2-3	<i>Häufigkeitsverteilung zur Lehrberufsdauer (in Jahren)</i>	XXVI
Tabelle C.2-4	<i>Häufigkeitsverteilung zur Dauer der Unterrichtstätigkeit in der erhobenen Klasse</i>	XXVI
Tabelle D.1-5	<i>Darstellung der deskriptiven Itemkennwerte zur Unterrichtswahrnehmung</i>	XXX
Tabelle D.1-6	<i>Angabe der Items und ihrer deskriptiven Kennwerte, die sich hinsichtlich ihrer Schiefe- und Exzesswerte im Vorzeichen unterscheiden</i>	XXXVI
Tabelle D.2-7	<i>Darstellung der Item- und Reliabilitätswerte zur Unterrichtswahrnehmung</i>	XXXVII
Tabelle D.3-8	<i>Faktorladungsmuster für die 5-Faktorenlösung (HKA, Oblimin) für die ich-/wir-Version und die Gesamtstichprobe</i>	XLIV
Tabelle D.3-9	<i>Überblick zu den Fitindizes</i>	L
Tabelle D.4-10	<i>1-F-ANOVA zur Abweichung der Schüler- und Lehrerangaben in Abhängigkeit vom Unterrichtsmerkmal</i>	LIII



## **Zusammenfassung**

Der Fokus dieser Arbeit liegt auf der Entwicklung und Überprüfung eines Fragebogens zur Erfassung der Unterrichtsqualität aus Schülersicht. Das Verfahren soll in der schulischen Selbstevaluation als Feedbackinstrument für Lehrkräfte im Rahmen der Unterrichtsentwicklung in Kombination mit einem inhaltsäquivalenten Fragebogen zur Selbsteinschätzung Anwendung finden.

Ausgangspunkt ist ein von Unterrichtsforschern konstruierter fächerübergreifender Itempool zu Kernmerkmalen der Unterrichtsqualität, der im Rahmen dieser Arbeit in einem mehrstufigen Verfahren überprüft wird. Im ersten Schritt bearbeiten Schüler den Fragebogen in qualitativen Einzelinterviews, anschließend wird der entsprechend veränderte Fragebogen in ca. 80 Klassen mit 1800 Schülern/innen in verschiedenen Fächern der Sekundarstufe I an rheinland-pfälzischen Schulen erprobt. Dabei kommen hinsichtlich der Wahrnehmungsperspektive unterschiedliche Versionen zum Einsatz. Die Lehrkräfte erhalten im Anschluss eine graphisch aufbereitete Rückmeldung der Ergebnisse.

Die Datenbasis wird anschließend ausführlichen testtheoretischen und faktorenanalytischen Verfahren unterzogen. Neben Reliabilitäts- und Validitätsaspekten werden explizit auch Kriterien der Handhabbarkeit und Praktikabilität in die Entscheidung zur Itemselektion einbezogen. Die integrative Bewertung aller Kriterien mündet schließlich in einen Fragebogen, der anhand von insgesamt 60 Items die Merkmale Klassenführung, Lernförderliches Klima, Motivierung, Klarheit und Verständlichkeit sowie Aktivierung aus der Perspektive des einzelnen Schülers erfasst.

### **Freie Schlagwörter (deutsch):**

Unterrichtsqualität – Schülerwahrnehmung – Fragebogen – Unterrichtsentwicklung – Selbstevaluation – Interne Evaluation

### **Freie Schlagwörter (englisch):**

quality of instruction – student ratings – questionnaire – instructional development – self-evaluation – internal evaluation

## 1 Einleitung

*„Das Wahre gibt es nicht!  
Es gibt nur verschiedene Arten des Sehens.“  
Gustave Flaubert (1821-1880)*

Die Prozesse der empirischen Wende (Lange, 2001) und die gewonnenen Erkenntnisse aus nationalen und internationalen Schulleistungstudien, wie z. B. PISA und TIMSS, haben im deutschen Bildungssystem zu einem Paradigmenwechsel vom input- zum ergebnis- bzw. outputorientierten Steuerungsmodell geführt. Qualitätssicherung wird dadurch zu einem festen Bestandteil des schulischen Aufgabenspektrums. Dies bringt neben einer gewissen pädagogischen und organisatorischen Gestaltungsfreiheit auch die Forderung nach einer hohen Qualitätsverantwortung mit sich. Diese Entwicklung und damit einhergehende gesetzliche Vorgaben resultieren in der Notwendigkeit interner und externer Evaluationsmaßnahmen. Die Steuerung der Überprüfungs- und Verbesserungsprozesse zur Qualität von Bildung und Unterricht wird dabei mehr und mehr auf die Einzelschulebene verlagert. Das Schlagwort ‚Selbstständigkeit bzw. Autonomie der Schule‘ kennzeichnet diesen Wandel. Hinsichtlich der Selbstevaluation stehen den Schulen jedoch kaum wissenschaftlich fundierte Instrumente zur Verfügung. Diese Lücke mit einem überprüften Schülerfragebogen als Feedbackinstrument zur Unterrichtsqualität zu schließen, ist das Ziel dieser Arbeit.

Die Schülerperspektive wird vor dem Hintergrund gewählt, dass die Schüler<sup>1</sup> als Adressaten des Unterrichts hinsichtlich der Wirkung pädagogischer Prozesse als Experten angesehen werden können. Hofer rät bereits 1981 Lehrkräften zum Einholen von Schülerfeedback: „Auf jeden Fall sucht eine reife, professionelle Persönlichkeit soviel wie möglich einschlägige Rückmeldung, um sich selbst bewerten zu können. Lehrer haben die beste Gelegenheit, sich über ihre Schüler unmittelbare Rückmeldung zu verschaffen. Sie sollten sie nutzen.“ (S. 54). Das systematische Einholen von Schülerfeedback ist allerdings auch 25 Jahre nach diesem Appell eher die Ausnahme als die Regel, obwohl die jüngsten Debatten bezüglich der Veröffentlichungen von Lehrerbewertungen im Internet zeigen, dass Schüler sich wünschen, in ihrer Meinung stärker wahrgenommen zu werden. Auf Lehrerseite führt der Schülerfeedbackansatz zu unterschiedlichen Reaktionen – von Ablehnung über Skepsis bis hin zur Akzeptanz und dem überzeugten Einsatz im Schulalltag. Die Art

---

<sup>1</sup> In der vorliegenden Arbeit wird aus Gründen der sprachlichen Vereinfachung auf geschlechtsspezifische Formulierungen verzichtet; die gewählte Form steht sowohl für das weibliche als auch männliche Geschlecht.

und Weise, wie Schülerfeedback eingeführt, durchgeführt und gelebt wird, sind medien- de Faktoren für diese Sichtweisen. Gelingt die Integration des Schülerfeedbacks in die Schulkultur und den pädagogischen Alltag der einzelnen Lehrkraft, profitieren sowohl die Schüler als auch die Lehrer. Bleibt der Einsatz des Schülerfeedbackverfahrens ein singulä- res Ereignis, können die Potentiale dieses Ansatzes nicht ausgeschöpft werden und drohen, als Eintagsfliege zu verpuffen. Ein weiteres Ziel dieser Arbeit ist es daher, konzeptionelle Überlegungen aufzuführen, wie die Implementierung des Schülerfeedbackgedankens in den pädagogischen Schulalltag gelingen kann.

## **1.1 Fragestellung und institutionelle Einbettung**

Die derzeitige Bildungsforschung ist geprägt durch den Fokus auf die schulische Evaluati- on mit der Intention, die Leistungsfähigkeit des deutschen Schulsystems zu erhöhen. Hin- sichtlich der Verbesserung der schulischen Qualität kommt dem Unterricht im komplexen schulischen Wirkgefüge eine Schlüsselrolle zu (Ditton, 2002a; Helmke, 2006b), weshalb „Reformen im schulischen und kontextuellen Rahmen...gezielt auf Verbesserungen des Unterrichts ausgerichtet sein“ (Ditton, 2002a, S. 262) sollten.

Ein Element zur Sicherstellung der Qualität pädagogischer Interaktionen im Unterricht ist die kontinuierliche Verbesserung der professionellen Handlungskompetenz von Lehrkräften (Baumert & Kunter, 2006). Eine Komponente dieser ist die Diagnosekompetenz, die neben der Analyse von Fähigkeitsunterschieden als Inputfaktor und Lernergebnissen von Schülern als Unterrichtsprodukt auch die Einschätzung von Unterrichtsprozessen umfasst, wobei letzterem in der Forschung bisher kaum Beachtung geschenkt wurde (Helmke, 2003). Studien weisen eine starke Diskrepanz zwischen der Selbsteinschätzung durch die Lehr- kraft und z. B. der per Videografie erhobenen Prozessmerkmale des Unterrichts auf (vgl. VERA: Helmke, Helmke, Heyne, Hosenfeld, Hosenfeld, Schrader & Wagner, 2008; vgl. DESI-Studie: Helmke, Helmke, Schrader, Wagner, Nold & Schröder, 2008). Um diese Lücke zu schließen, widmet sich das Gemeinschaftsforschungsprojekt<sup>2</sup> „Unterrichtsdia- gnostik: Beurteilung des Unterrichts als Bestandteil diagnostischer Kompetenz und pädä- gogischer Professionalität“ z. B. den Fragestellungen, wie gut Lehrkräfte Unterrichtsmerk- male einschätzen können und welche Rolle Urteilsvoreingenommenheiten, subjektive

---

<sup>2</sup> Das Kooperationsprojekt wird vom Hochschulsonderprogramm „Wissen schafft Zukunft“ der Landesregie- rung Rheinland-Pfalz finanziert und vom Ministeriums für Bildung, Wissenschaft, Jugend und Kultur Rhein- land-Pfalz (MBWJK) nachhaltig unterstützt.

Unterrichtstheorien sowie der persönliche Unterrichtsbezug spielen (Helmke, Helmke, Schrader, Wagner, Kleinbub & Pikowsky, 2007). Das Projekt zielt ferner auf die Entwicklung wissenschaftlich fundierter und gleichzeitig praktikabler Instrumente zur Unterrichtsdiagnostik und -entwicklung im internen Evaluationskontext ab. Denn u. a. besagt das Schulgesetz von Rheinland-Pfalz, dass Schulen nach dem Festlegen pädagogischer Schwerpunkte und Ziele zur Sicherung und Entwicklung der schulischen Qualitätsarbeit neben der Teilnahme an externen Evaluationsmaßnahmen dazu verpflichtet sind „regelmäßig das Erreichen dieser Ziele (interne Evaluation) [zu überprüfen]“ (Staatskanzlei, 2004, Teil 2, Abschnitt 1, §23). Auch die OECD versteht die regelmäßige Reflexion der eigenen Leistungen und Expertise als ein wesentliches Element professioneller Kompetenzen (OECD, 1989) – allerdings mangelt es diesbezüglich an adäquaten überprüften Instrumenten zur internen Evaluation.

In diesem Bereich ist das Diplomarbeitsvorhaben als Teilprojekt angesiedelt, denn dieses fokussiert die Weiterentwicklung und Überprüfung eines fragebogenbasierten Diagnostikverfahrens zur Unterrichtsqualität aus der Schülerperspektive. Der Fragebogen soll als internes Evaluations- und Feedbackinstrument zur Unterrichtsentwicklung sowie zur Professionalisierung der pädagogischen Tätigkeit beitragen, indem es die Wirksamkeit der Lehr-Lernprozesse überprüft und Steuerungswissen im Kontext der Selbstevaluation verfügbar macht. Ditton und Arnoldt (2004) betonen, dass der zuletzt genannte Prozess in seiner Schwierigkeit nicht unterschätzt werden darf, da Veränderungen in der Regel zu Beginn auf Widerstand stoßen. Deshalb widmet sich diese Arbeit auch den Fragen, wie der Fragebogen konkret im Schulalltag angewendet werden und wie die Etablierung des Schülerfeedbacks gelingen kann.

Die Arbeit basiert, wie in Abbildung 1 veranschaulicht, auf einem forschungsbezogenen Kooperationsprojekt zwischen der Universität Landau, dem Institut für schulische Fortbildung und schulpsychologische Beratung des Landes Rheinland-Pfalz (IFB) und der Universität Mannheim. Der überprüfte Fragebogen wird nach Projektabschluss interessierten Lehrkräften durch das IFB in seiner Funktion, Lehrkräften und Schulen mit dem Fokus auf den jeweiligen schulischen Bedarf vor Ort und die bildungspolitischen Zielsetzungen ein umfassendes Weiter-, Fortbildungs-, Beratungs- und Methodenangebot zur Verfügung zu stellen, angeboten.

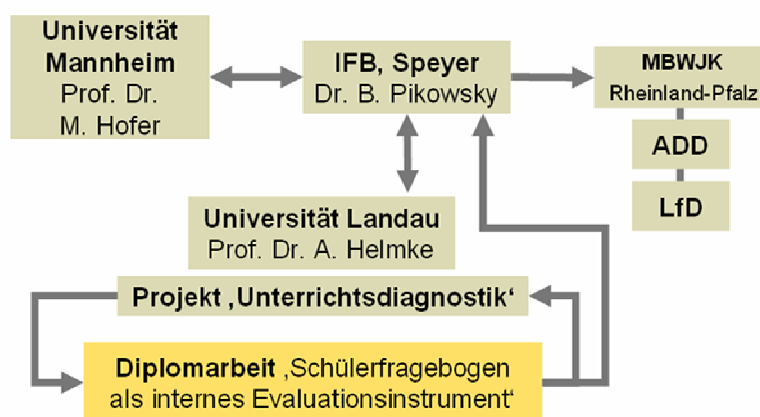


Abbildung 1: Darstellung der institutionellen Einbettung dieser Arbeit

## 1.2 Zielsetzung und Aufbau

Das Ziel der Diplomarbeit ist die Überprüfung eines Fragebogenverfahrens, welches die Unterrichtsqualität aus der Schülersicht erfasst und als Selbstevaluationsinstrument in der Sekundarstufe I zur Anwendung kommen soll. Die sich aus dieser Fragestellung ergebenden Zielsetzungen der Arbeit sind: (1) Die Datenerhebung und -aufbereitung zur Überprüfung des Itempools sowie die anschließende Auswertung des Schülerdatensatzes, um eine Entscheidung hinsichtlich der in den endgültigen Fragebogen aufzunehmenden Items treffen zu können und diesen zu erstellen. (2) Die Erörterung der Fragestellung, wie der Fragebogen als Feedbackverfahren in den pädagogischen Alltag der einzelnen Lehrkraft aber auch als Element der Schulentwicklung insgesamt integriert werden kann. Der zu überprüfende Itempool umfasst 78 Items, die von einem Expertenteam<sup>3</sup> der Unterrichtsforschung formuliert wurden und fünf Merkmale guten Unterrichts gemäß Helmke (2007c) erfassen: Klassenführung (17 Items), Lernförderliches Klima (14 Items), Motivierung (13 Items), Klarheit und Verständlichkeit (16 Items) und Aktivierung (18 Items). Bei der inhaltlichen Itemformulierung griff das Expertenteam maßgeblich auf den von Helmke entwickelten Unterrichtsbeobachtungsbogen (UBB, Helmke, 2007b) und den darauf basierenden und von Helmke gemeinsam mit dem wissenschaftlichen Beirat der AQS<sup>4</sup> entwickelten Beobachtungsbogen ‚Einblicknahmen in die Lehr- und Lernsituationen‘ (Helmke & AQS, 2007) zurück, der in der externen Evaluation zum Einsatz kommt. Außerdem fließen Items

<sup>3</sup> Das Expertenteam besteht maßgeblich aus Prof. Andreas Helmke, Dr. Friedrich-Wilhelm Schrader und Dr. Wolfgang Wagner der Universität Koblenz-Landau.

<sup>4</sup> Die Agentur für Qualitätssicherung, Evaluation und Selbstständigkeit von Schulen (AQS) ist eine Serviceeinrichtung des Landes Rheinland-Pfalz, deren Auftrag die Durchführung externer Schulevaluationen ist.

der im Rahmen von IQES online<sup>5</sup> angebotenen, und ebenfalls von Helmke unter Mitarbeit von Brägger und Wagner entwickelten, Instrumente zur Erfassung der Qualität des Mathematik-, Deutsch- und Englischunterrichts aus der Schülerperspektive ein. Die Items sind fächerübergreifend auf der Ebene konkret beobachtbaren Verhaltens formuliert und lassen sich sowohl aus der Schüler- als auch der Lehrerperspektive in äquivalenter Formulierung beurteilen. Hinsichtlich der Wahrnehmungsperspektive gibt es zwei Versionen (ich- vs. wir-Bezug), die bei der Befragung innerhalb der Klasse nach dem Prinzip odd & even rotiert werden. Um eine Generalisierung über verschiedene Fächer, Lehrkräfte und die Zeit zu vermeiden, soll der Fragebogen in einem bestimmten Fach zur Beurteilung einer konkreten Unterrichtsstunde eingesetzt werden. Neben der Unterrichtseinschätzung wird auch die subjektive Sicherheit dieses Urteils erfragt.

Die vorliegende Arbeit ist so aufgebaut, dass zunächst der theoretische und empirische Wissensstand zum Fragebogenkonstrukt Unterrichtsqualität dargestellt (Kapitel 2) und anschließend die Unterrichtswahrnehmung aus Schülersicht näher betrachtet wird (Kapitel 3). Die Frage, welchen Beitrag Schülerfeedback als Selbstevaluationsinstrument zur Unterrichts- und Schulentwicklung leisten kann, ist Gegenstand des Kapitels 4. Der Theorieteil schließt mit der Präzisierung der Zielsetzungen dieser Arbeit.

Der empirische Teil stellt die einzelnen Schritte des mehrstufigen Itemüberprüfungsverfahrens dar. Von der qualitativen Voruntersuchung (Kapitel 6) ausgehend, wird nach der Beschreibung der Haupterhebung (Kapitel 7) sowie der empirischen Vorarbeiten (Kapitel 8) die deskriptivstatistische, testtheoretische und faktorenanalytische Untersuchung der Items zur Unterrichtswahrnehmung dargestellt (Kapitel 9). Die Angaben zur Urteilssicherheit werden im Kapitel 10 analysiert. Die Zusammenführung der Analyseergebnisse hinsichtlich einer endgültigen Itemselektion wird in Kapitel 11 vollzogen. Die Arbeit schließt mit einer kritischen Reflexion der verwendeten Methodik, einer Bewertung des überprüften Fragebogens hinsichtlich relevanter Gütekriterien sowie weiterführenden Forschungs- und Anwendungsfragen.

---

<sup>5</sup> IQES online ([www.iqesonline.net](http://www.iqesonline.net)) bietet als Web-Plattform "Instrumente für die Qualitätsentwicklung und Evaluation in Schulen".

## **I Theoretischer Teil**

Die Unterrichtsqualität als das von den Schülern und Lehrkräften zu bewertende Fragebogenkonstrukt ist in der vorliegenden Arbeit von zentraler Bedeutung. Daher wird als theoretischer Hintergrund zunächst ein Überblick über wesentliche Forschungsbefunde zur Unterrichtsqualität gegeben (Kapitel 2). Anschließend werden relevante Aspekte zur Wahrnehmung und Erhebung der Unterrichtsqualität aus der Schülerperspektive erläutert (Kapitel 3). Im Kapitel 4 wird das Schülerfeedbackverfahren in seiner Funktion als Selbstevaluationsinstrument im Kontext der Unterrichtsentwicklung betrachtet und ein Rahmenmodell zur Implementierung des Fragebogenverfahrens als pädagogisches Feedbackinstrument für Lehrkräfte vorgestellt. Der theoretische Teil schließt mit der Konkretisierung der Fragestellungen der vorliegenden Arbeit im Kapitel 5.

## **2 Unterrichtsqualität als Meta-Konstrukt**

Nach allgemeinen Ausführungen zur Konzeptualisierung der Unterrichtsqualität (Abschnitt 2.1), nähert sich dieses Kapitel über ausgewählte Modellvorstellungen und Befunde der Lehr-Lernforschung (Abschnitt 2.2) der Antwort zur Frage „Was ist guter Unterricht?“. Anschließend erfolgt die Darstellung variablenzentrierter Strukturierungsansätze prozessualer Unterrichtsqualitätsdimensionen als Grundlage für die Fragebogenkonstruktion (Abschnitt 2.3).

### **2.1 Konzeptualisierung von Unterrichtsqualität**

Auf modelltheoretischer Ebene verwendet Carroll gemäß Einsiedler (1997) als einer der ersten den Begriff Unterrichtsqualität in seinem Modell schulischen Lernens von 1964 (s. Abschnitt 2.2.1). 25 Jahre später definieren Weinert, Schrader und Helmke (1989) das Metakonstrukt Unterrichtsqualität als „jedes stabile Muster von Instruktionsverhalten, das als Ganzes oder durch einzelne Komponenten die substantielle Vorhersage und/oder Erklärung von Schulleistung erlaubt“ (S. 899, übersetzt nach Einsiedler, 1997, S. 228). Einsiedler definiert Unterrichtsqualität 2002 als „...Bündel von Unterrichtsmerkmalen, die sich als „Bedingungsseite“ (oder Prozessqualität) auf Unterrichts- und Erziehungsziele („Kriterienseite“ oder Produktqualität) positiv auswirken, wobei die Kriterienseite überwiegend von normativen Festlegungen bestimmt ist und der Zusammenhang von Unterrichtsmerkmalen und Zielerreichung von empirischen Aussagen geleitet ist“ (S. 195). In dieser Zeitspanne

widmen sich unterschiedliche Forschungsansätze und -traditionen der Frage, durch welche bedeutsamen Merkmale und Komponenten sich die Lernumwelt eines ‚guten Unterrichts‘ beschreiben und erfassen lässt. Ditton (2002b) gruppiert diese in die Lehr-Lernforschung, die Schul- bzw. Klassenklimaforschung, die Schulqualitätsforschung und Unterrichtsmodelle der allgemeinen bzw. fachspezifischen Didaktik. Letztere fokussiert fachdidaktische Reflexionen der intendierten und vorhandenen Praxis unter einem Ziel-Inhalt-Methodenbezug (Berner, 1999) und wird in den folgenden Betrachtungen ausgeklammert<sup>6</sup>. Die anderen Forschungsrichtungen beschreiben die Lehr-Lernprozesse auf der Mikro- (z. B. Unterricht) bzw. Makroebene (z. B. Schulqualität, vgl. Schnabel, 2001) und werden im Folgenden überblicksartig skizziert.

Die *Lehr-Lernforschung*, auch Unterrichtsforschung oder Instruktionsforschung genannt, entwickelte sich in den 60er Jahren und fokussiert die theoriegeleitete Beschreibung, Erklärung und Optimierung von Lehr- und Lernprozessen insbesondere unter den Bedingungen institutionalisierten Unterrichts (Weinert & Treiber, 1982). Bei dieser Mikro-Ebenen-Perspektive wird das Lehrerverhalten in der Lehrer-Schüler-Interaktion mit Konzepten wie ‚Lehrereffektivität‘ (teacher effectiveness) und ‚Klassenführung‘ in den Vordergrund der Betrachtung gestellt. Die ersten Modelle zur Unterrichtsqualität beruhen im Kern auf der Unterscheidung der Unterrichtsqualität von der zur Verfügung stehenden Lernzeit (Quantität) und tragen deshalb den Namen *Lehr-Lernzeit-Modelle* (Einsiedler, 2002; Gruehn, 2000, s. dazu auch Abschnitt 2.2.1). Die Lehr-Lernforschung beruht dabei auf einer möglichst objektiven Erfassung der Unterrichtsstrukturen und versucht, eine kausale Verbindung zum Lernerfolg und Verhalten des Schülers abzuleiten. Die Größe des Einflusses der Unterrichtsqualität auf die Schulleistung und Entwicklung der Schüler ist auch bezüglich der Abschätzung der Effektivität des Fragebogens als pädagogische Intervention zur Unterrichtsentwicklung eine wichtige Frage. Deshalb soll zunächst ein Blick auf Befunde zu dieser Fragestellung geworfen werden.

Der Coleman-Report (Coleman, Campbell, Hobson, McPartland, Mood, Weinfeld & York, 1966) und die Veröffentlichung von Jencks und Mitarbeitern (Jencks, Smith, Acland, Bane, Cohen, Gintis, Heyns & Michelson, 1972) regten eine heftige Diskussion über die Unterrichtswirksamkeit und auch die der Schule im Allgemeinen bezüglich der Leistungsentwicklung der Schüler an. Ihre Forschungsergebnisse zeigen einen durch die Schulzugehö-

---

<sup>6</sup> Weiterführende Informationen zu didaktischen Modellen findet der interessierte Leser bei Ditton (2002b).



rigkeit aufgeklärten Varianzanteil der Leistungsunterschiede von Schülern von 1-3% (Jencks et al., 1972) bzw. maximal 1% (Coleman et al., 1966). Rutter, Maughan, Mortimore und Ouston (1979) führten allerdings eine Reihe methodischer Einwände ins Feld und zeigten mit einem längsschnittlichen Vorgehen, dass bei der Verwendung von Leistungstests als Outputvariable (im Gegensatz z. B. zur verbalen Ausdrucksfähigkeit bei Coleman) und der Operationalisierung der Unterrichtsqualität auf der Prozessebene (im Gegensatz zur z. B. Bücheranzahl oder der Lehrer-Schüler-Relation als Schulvariablen bei Coleman), der aufgeklärte Varianzanteil bei Konstanthaltung der kognitiven Eingangsvoraussetzungen ca. 10% beträgt. Gruehn (2000) merkt des Weiteren kritisch an, dass die Höhe des auf die Schulzugehörigkeit zurückführbaren Varianzanteils stark vom Ausmaß der Leistungs differenzierung des entsprechenden Schulsystems abhängt. Ca. 2,5% der Schulleistungsvarianz werden z. B. im deutschen Schulsystem nach Ditton und Kreckler (1995) auf die Schulzugehörigkeit innerhalb einer Schulart zurückgeführt. Wird auf den Leistungsvarianzanteil zwischen den einzelnen Schulformen rekurriert, erhöht sich dieser auf 27% bis 43%. In einem Einheitsschulsystem, wie z. B. dem US-amerikanischen, wären diese Varianzanteile innerhalb und nicht zwischen den Schulen verortet, denn diese differenzieren über ein entsprechendes Kursangebot innerhalb der einzelnen Schule (Gruehn, 2000). Deshalb kann an die Stelle einer pessimistischen Skepsis durchaus die optimistische Haltung gesetzt werden, dass die Unterrichtsqualität nicht nur die Lernleistungen der einzelnen Schüler verbessert, sondern gemäß Gruehn (1995) sogar die motivationale und affektive Entwicklung der Schüler positiv beeinflusst (s. dazu auch Abschnitt 2.2.4). Werden diese Befunde vor dem Hintergrund des Fragebogeneinsatzes betrachtet, lässt sich festhalten, dass dadurch *eine* notwendige Voraussetzung zur Fragebogeneffektivität erfüllt ist (s. dazu auch Abschnitt 4.2).

Im Gegensatz zur Lehr-Lernforschung beginnen Ende der 70er Jahre die *Schul- und Klassenklimaforscher* – angeregt durch die ‚kognitive Wende‘ – verstärkt die subjektiven Empfindungen und Wahrnehmungen der Schüler als mediiierende Variable zu reflektieren (Achtenhagen, Sembill & Steinhoff, 1979; Kahl, 1977; Kahl, Buchmann & Witte, 1977). Hintergrund ist die Annahme, dass nicht die objektiven Umweltgegebenheiten, sondern die subjektive Wahrnehmung dieser das Verhalten und die Entwicklung des Einzelnen bestimmen (Pekrun, 1985). Das subjektive Erleben innerschulischer Umwelten wird dabei sowohl als abhängige als auch als unabhängige Variable hinsichtlich unterrichtlicher und schulischer Kriterien zum Untersuchungsgegenstand (Ditton, 2002b). Eine positiv erlebte schuli-

sche Umwelt trägt u. a. zum Anstieg schulspezifischer Interessen (Eder, 1992), zu einer intensiveren Mitarbeit im Unterricht (Fend, 1977) sowie einer positiveren Ausprägung des leistungsspezifischen Selbstkonzepts (Pekrun, 1985) bei und erhöht die Zufriedenheit der Schüler mit der Schule (vgl. Fend, 1977; Lange, Kuffner & Schwarzer, 1983). Die Vorhersageleistung von Klimadimensionen (u. a. Lehrer-Schüler-, Schüler-Schüler-Beziehungen aber auch Unterrichtsmerkmale, vgl. Gruehn, 2000, S. 73) hinsichtlich schulischer Leistungskriterien ist niedrig, aber gemäß Eder (2002) zumindest in Studien zum amerikanischen Schulsystem konsistent positiv. Gruehn (2000) kann allerdings in ihrer Studie für die erfassten affektiven Qualitätsmerkmale der Lehrer-Schüler-Beziehung, wie z. B. der Sozialorientierung des Fachlehrers und die Zufriedenheit mit diesem keine signifikanten Zusammenhänge mit der über standardisierte Leistungstests operationalisierten Leistungsentwicklung finden.

Sowohl im Forschungsprogramm der Lehr-Lernforschung als auch der Klimaforschung werden gesellschaftliche Rahmenbedingungen, lokale oder auch soziale Bildungskontexte sowie außerschulische Medien und schulische Organisationsformen in der Regel vernachlässigt (Weinert & Treiber, 1982). Diese Lücke schließt die *Schulqualitätsforschung*, indem die Makro-Ebene als Betrachtungsperspektive gewählt wird. Die Schulqualitätsforschung sucht in erster Linie nach Merkmalen guter Schulen im Allgemeinen und nähert sich darüber auch relevanten Kernmerkmalen guten Unterrichts (Ditton & Kreckler, 1995). Den Anstoß dazu gibt der Befund, dass schulische Qualitätsfaktoren keine konstante prädiktive Validität hinsichtlich der Schülerleistungen aufweisen, sondern in Abhängigkeit von der Zeit oder z. B. dem Unterrichtsfach variieren. Der Rückschluss auf vorhandene Lehrkraft- und Unterrichtsunterschiede liegt nahe, wenn von einer ähnlichen Schülerzusammensetzung innerhalb der einzelnen Schulen ausgegangen wird (Luyten, 2003; Teddlie & Reynolds, 2001). Als Beispiel für Modelle zur Schul- und Bildungsqualität seien hier folgende genannt: Dittons Modell zur Bildungsqualität (Ditton, 2000), Integriertes Modell der Schulleffektivität (Scheerens, 1992) und das MACRO-Modell (Stringfield, 1994).

Innerhalb der verschiedenen Forschungsansätze ist die Konzeptualisierung der Unterrichtsqualität wiederum vielschichtig. Die Anforderungen an den Lehrerberuf werden auch im Laufe des gesellschaftlichen Wandels neu definiert. Dies schlägt sich u. a. in neuen Richtungen innerhalb der empirischen Unterrichts- und Lehrerforschung nieder. Zuerst liegt der Fokus des sogenannten *Persönlichkeitsparadigmas* aus einer differentiellen Perspektive auf

den Persönlichkeitsmerkmalen der Lehrkraft als Prädiktoren für fachliche Leistungen, aber auch für Kriterien wie z. B. die Angstfreiheit der Schüler. Ein generelles Persönlichkeitsprofil eines ‚guten Lehrers‘ lässt sich dabei jedoch nicht identifizieren, allenfalls Mindestausprägungen (Krapp & Weidenmann, 2001). Einen geringeren Abstraktionsgrad weist das sich später etablierende behavioristisch geprägte *Prozess-Produkt-Paradigma* auf, welches aus einer verhaltensorientierten Perspektive einzelne Lehrerverhaltensmuster (Prozessmerkmale) hinsichtlich ihrer unmittelbaren oder längerfristigen Effekte auf Schülerseite (Produktmerkmale) untersucht (Gruehn, 2000; Niegemann, 2001). Wesentliche Kritikpunkte sind (1) die Vernachlässigung von Interaktionseffekten, die u. U. zu gegenläufigen Wirkungen der gleichen Instruktionsmaßnahmen bei unterschiedlichen Schülern führen können (Aptitude-Treatment-Interaktion, ATI, Snow & Swanson, 1992). (2) Die ausschließliche Konzentration auf die Wirkrichtung von der Lehrkraft zum Schüler, denn auch das Schülerverhalten beeinflusst das Verhalten der Lehrkraft (Krapp & Weidenmann, 2001) und (3) die Vernachlässigung der Merkmale der Lehrerpersönlichkeit (Helmke, 2007c). Durch eine Hinwendung zu Mediationsprozessen zwischen dem Unterrichtshandeln der Lehrkraft und dem Schülerverhalten werden diese Kritikpunkte in einer späteren Phase des Prozess-Produkt-Paradigmas aufgeweicht. Soziale Verhaltensmuster innerhalb der Klasse und die Informationsverarbeitungsprozesse der Schüler als aktive Lerner treten unter einer ‚Prozess-Prozess-Produkt‘-Fokussierung stärker in den Mittelpunkt (vgl. Gruehn, 2000). Im kognitionspsychologisch orientierten *Expertenparadigma* wird nach einem Ensemble an Kognitionen und Professionswissen gesucht, welches die Bewältigung der Anforderungen, die an die Lehrkraft im beruflichen Alltag gestellt werden, ermöglicht. Experten-Novizen-Vergleiche zeigen, dass die Lehrkräfte als Unterrichtsexperten den Unterricht in Episoden wahrnehmen, die der gemeinschaftlichen Inszenierung des Stoffes dienen und situationsangemessen elaborierte Handlungsziele verfolgen (vgl. Gruehn, 2000; Krapp & Weidenmann, 2001).

Im Folgenden werden nun ausgewählte Modelle der Unterrichtsforschung hinsichtlich ihrer Konzeptualisierung der Unterrichtsqualität vorgestellt.

## 2.2 Modelle und Befunde der Unterrichtsforschung

In den ersten Abschnitten (2.2.1-2.2.3) werden vor dem Hintergrund bedeutsamer Lehr-Lernzeit-Modelle wesentliche Annahmen und Befunde bezüglich der Unterrichtsqualität aus der Sicht der Lehr-Lernforschung aufgezeigt. Darauf folgt die Darstellung des Angebot-Nutzungs-Modells von Helmke, welches als theoretischer Rahmen des zu überprüfenden Fragebogenverfahrens fungiert (Abschnitt 2.2.4).

### 2.2.1 Carrolls und Blooms Modell schulischen Lernens

Carroll (1964) postuliert in seinem *Modell schulischen Lernens* (s. Abbildung 2) den Grad des Lernerfolgs i. S. des individuellen und kollektiven Schulleistungsniveaus als eine Funktion der tatsächlich aufgewendeten Lernzeit in Relation zur benötigten Lernzeit. Die benötigte Lernzeit wird dabei durch die aufgabenspezifische Begabung und die Fähigkeit, dem Unterricht zu folgen, welche wiederum durch die kognitiven Eingangsvoraussetzungen des Schülers und die Unterrichtsqualität beeinflusst werden, bestimmt. Die Ausdauer des Schülers und die ihm zugestandene Lernzeit bedingen die tatsächlich aufgewendete Lernzeit. Die Unterrichtsqualität mediiert dabei mit Merkmalen wie Klarheit, Adaptivität, Sequenzierung und Berücksichtigung der Schülerbedürfnisse über die Ausdauervariable und die Fähigkeit zum Instruktionsverständnis den Lernerfolg.

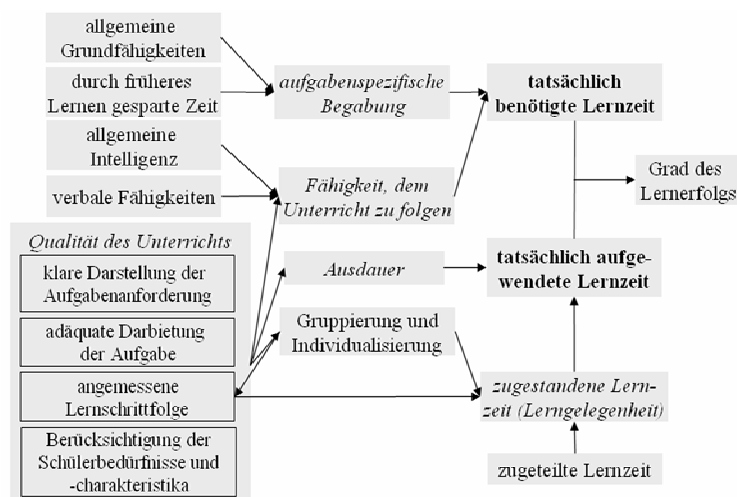


Abbildung 2: Carrolls Modell schulischen Lernens (1964) in Anlehnung an Harnischfeger und Wiley (1977, S. 209)

Im Rahmen des Modells wird postuliert, dass hohe Ausprägungen der Merkmale lernerfolgsförderlichen Unterrichts niedrige kognitive Eingangsvoraussetzungen kompensieren können, denn in diesem Fall ist die benötigte Lernzeit kaum von der Fähigkeit, dem Unter-

richt zu folgen, sondern maßgeblich von der aufgabenspezifischen Begabung abhängig. Je niedriger hingegen die Merkmale der Unterrichtsqualität ausgeprägt sind, desto wichtiger werden im Umkehrschluss die Eingangsfähigkeiten des Schülers, da diese die Fähigkeit determinieren, dem Unterricht zu folgen. Carrolls Modell (1964) impliziert auch, dass bei Konstanthaltung der zugestandenen Lernzeit die aufgabenspezifische Begabung und somit die kognitiven Eingangsvoraussetzungen den Lernerfolg maßgeblich determinieren. Dies entspricht der schulischen Realität in Deutschland, denn über Curricula und Rahmenvorgaben wird festgelegt, innerhalb welchen Zeitraums die Lernaufgaben zu bewältigen sind. Variiert die zugestandene Lernzeit interpersonal, erhöht sich der Einfluss und somit die Bedeutung der Unterrichtsqualität als die entscheidende Wirkvariable bezüglich der zugestandenen und tatsächlich aufgewendeten Lernzeit.

Genau wie Carroll definiert Bloom (1976) die Qualität des Unterrichts als eine von den Eingangsfähigkeiten der einzelnen Schüler unabhängige Variable. Die Forschungsergebnisse von Weinert, Schrader und Helmke (1989) zeigen jedoch, dass dieses Postulat empirisch nicht haltbar ist, denn Lehrkräfte orientieren sich in ihrem Verhalten an den Schülervoraussetzungen und -merkmalen. Ferner haben Helmke und Weinert (1997, S. 145) gezeigt, dass das fachspezifische Vorkenntnisniveau zu Beginn eines Schuljahres 16% der Zeitnutzungsvarianz im Unterricht und 14% der Varianz des Unterrichtsmerkmals Klarheit erklärt. Weiterhin erachtet Bloom (1976) die Lernleistung, welche in seinem *Modell schulischen Lernens* über die Leistungshöhe und Leistungsart, die Lernrate und affektive Lernergebnisse (z. B. Zutrauen in die eigene Leistungsfähigkeit) operationalisiert wird, als Resultat kumulativer Lernprozesse: Die Bewältigung einer Lernaufgabe führt zu entsprechenden affektiv-motivationalen Lernergebnissen, die wiederum die Eingangsvoraussetzungen für die nächste Lernaufgabe sind. Die entscheidenden Prädiktoren der Lernergebnisse sind für ihn die Unterrichtsqualität sowie die kognitiven und affektiven Eingangsvoraussetzungen des Schülers; die Lernzeit spielt eine untergeordnete Rolle. Als Aspekte der Unterrichtsqualität erachtet er Strukturierungshinweise der Lehrkraft bezüglich der Lernstoffabfolge, Bekräftigungen durch Lob bzw. Tadel, eine aktive Schülerbeteiligung durch Management- und Monitoring-Strategien und Korrekturen durch Anregungen bzw. Rückmeldungen an die Schüler.

Die beiden Modelle stellen sehr einflussreiche Versuche zur Erklärung schulischen Lernens dar, so bildet Carrolls Modell z. B. gemäß Bloom (1968) die theoretische Basis für das

Konzept des zielerreichenden Lernens (*mastery learning*). Eine kritische Betrachtung zeigt aber, dass das Konstrukt Unterrichtsqualität nicht genau definiert ist (Niegemann, 2001). Bei Carroll bleibt ferner offen, was eine tatsächlich aufgewendete Lernzeit bzw. eine aktive Lernzeit (Treiber, 1982) bezüglich der Elaboration und Bewältigung der Lerninhalte im Einzelfall bedeutet. Der Einfluss der Lehrkraft kommt bei Carroll außerdem nur indirekt über die Wirkung der zugestandenen Lernzeit und die Unterrichtsqualität auf die individuellen Variablen der tatsächlich benötigten und aufgewendeten Lernzeit zur Geltung.

### **2.2.2 Weiterentwicklungen durch Creemers und Slavin**

Ausgehend von Carrolls Modell schulischen Lernens fokussiert Creemers (1994) in seiner ‚Theory of Educational Effectiveness‘ die differenzierte Aufschlüsselung des Konstrukts Unterrichtsqualität. Auf einer breiten empirischen Forschungsbasis postuliert Creemers drei Komponenten der Unterrichtsqualität, die sinnvoll kombiniert und gegenseitig aufeinander abgestimmt werden müssen, um von Synergieeffekten hinsichtlich des schulischen Lernens profitieren zu können: (1) Beschaffenheit der Lehrpläne, (2) innere Differenzierungsmaßnahmen und (3) Verhaltensmerkmale des Lehrers. Unter (1) fasst Creemers die Quantität des Lernstoffs bzw. des Unterrichts an sich sowie die Strukturierung und Klarheit des Lernstoffs, das Anwenden von Advance Organizers und das Durchführen von Evaluationen bzw. Feedbackmaßnahmen. Zu den inneren Differenzierungsmaßnahmen (2) gehören die adäquate Anwendung von Mastery Learning-Konzepten, die Bildung leistungshomogener Gruppen und der Einsatz kooperativer Lernformen. Als Lehrermerkmale (3) gibt Creemers schließlich ein effizientes Klassenmanagement, die Vergabe von Hausaufgaben adäquaten Inhalts und Umfangs, das Vermitteln eindeutiger Zielsetzungen, das Kommunizieren angemessener Leistungserwartungen, das Strukturieren und die klare Darstellung der Lerninhalte, das Überprüfen der Lernzielerreichung sowie gezielte Lernstandsrückmeldungen an die Schüler und bei Bedarf das Durchführen korrigierender Maßnahmen an. Creemers verweist anhand der Thematisierung der Lehrplanqualität auf den institutionellen Kontext bei gleichzeitiger differenzierter Analyse des Lehrerverhaltens.

Auch Slavin trägt zur Weiterentwicklung der Modelle schulischen Lernens bei. Mit seinem ‚Model of Instructional Effectiveness‘ (1994, s. Abbildung 3) postuliert er die multiplikative Verknüpfung der vier Faktoren (1) Unterrichtsqualität, (2) angemessenes Instruktionslevel, (3) Anreizwert und (4) Lernzeit hinsichtlich der Schulleistung als notwendige Voraussetzungen für einen effektiven Unterricht. Die sich in der englischen Bezeichnung aufgrund

der Anfangsbuchstaben das Akronym QAIT ergebenden vier Komponenten, leitet Slavin, genau wie Creemers aus den Ergebnissen empirischer Forschungsarbeiten ab.

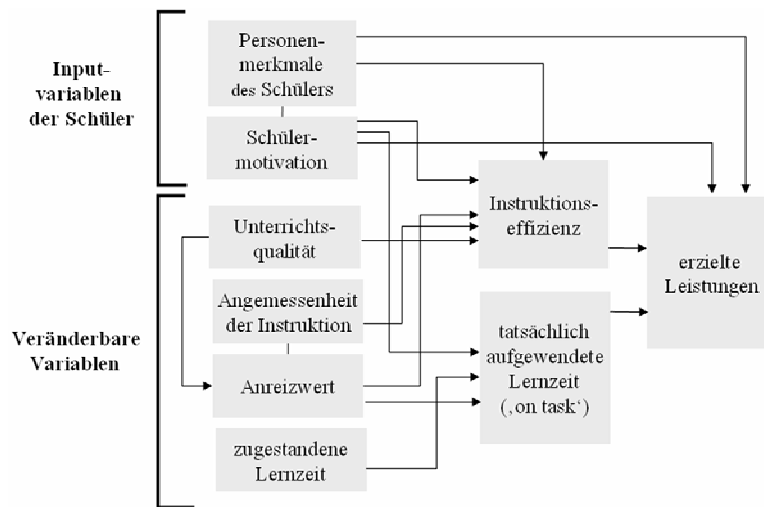


Abbildung 3: Slavins QAIT-Modell (in Anlehnung an Slavin, 1994, S. 149)

Die Unterrichtsqualität (1) sieht Slavin im Wesentlichen durch die Lehrplanqualität und eine klare, strukturierte und bildhafte Darstellung der Lerninhalte unter Bezug auf das Schülervorwissen sowie einen adäquaten Medieneinsatz bei einer enthusiastischen und humorvollen Art der Lehrkraft bedingt. Eine auf die verschiedenen Bedürfnisse der Schüler aufgrund eines unterschiedlichen Vorwissens und unterschiedlicher Lernraten abgestimmte Instruktion (2) sichert die Angemessenheit dieser. Empirisch zeigt sich, dass die effektivste Methode eine heterogen zusammengesetzte Klasse mit der temporären Bildung fähigkeits- bzw. leistungshomogener Klassen oder Subgruppen ist. Die konsequenteste Methode, um dies zu erreichen, wäre gemäß Slavin eine individualisierte Instruktion, welche allerdings Schwierigkeiten bezüglich des Klassenmanagements mit sich bringt und nicht durchweg positive Schulleistungseffekte zeigt (Hartley, 1977, zitiert nach Slavin, 1994). Extrinsische Anreize (3), wie z. B. Auszeichnungen oder Lob, können die Schüler motivieren, Lerngelegenheiten wahrzunehmen. Schüler können aber auch durch ansprechendes und interessantes Lehrmaterial intrinsisch zum Lernen motiviert werden. Gruehn (2000) weist darauf hin, dass gemäß Deci und Ryan (1985) auch Lob intrinsisch motiviertes Lernen bewirken kann. Die Lernzeit (4) ist in Anlehnung an Carrolls Modell abhängig von der zugestandenen Lernzeit und der tatsächlichen Lernzeit ‚on task‘. Erstere kann durch die Lehrkraft direkt reguliert werden, wohingegen die tatsächliche Lernzeit ein Produkt der Lerngelegenheiten, der Schülermotivation und der Unterrichtsqualität ist. Dies schlägt sich im QAIT-Modell in der Darstellung der tatsächlichen Lernzeit als medierende Variable und der zugestandenen

Zeit als veränderbare exogene Variable nieder. Die Modellstruktur verdeutlicht gleichzeitig einen Kernpunkt der Ausführungen von Slavin: „...effective instruction is not just good teaching“ (Slavin, 1994, S. 143). Eine notwendige Bedingung für einen effektiven Unterricht ist, dass alle vier Faktoren auf den jeweiligen Kontext zugeschnitten sind. Einen hinreichenden Charakter bekommt die Modellkonstellation erst, wenn die Schüler zusätzlich über ein entsprechendes Vorwissen und Motivation verfügen und sie die benötigte Lernzeit auch tatsächlich aufbringen.

Slavin und Creemers weisen eine deutliche Übereinstimmung in ihren konstatierten Unterrichtsmerkmalen und auch ihrer empirisch fundierten Herangehensweise auf. Dabei ist anzuerkennen, dass ihnen bei ihren Modellausführungen die Integration und Systematisierung bisheriger Befunde der Unterrichtsforschung gelungen ist. Außerdem verdeutlichen sie, dass Unterricht ein „komplexes Gefüge mit mehr oder weniger effektiven Kombinationen unterschiedlicher Unterrichtsmethoden und Lehrerverhaltensweisen“ ist (Gruehn, 2000, S. 12) und erreichen dadurch eine Umfokussierung weg von der Vorstellung, dass Unterricht ein aus singulären Komponenten bestehendes Konstrukt darstellt.

### 2.2.3 Walbergs Produktivitätsmodell

Als Synthese der Modelle schulischen Lernens kann *Walbergs Produktivitätsmodell* (Walberg, 1981; 1990) gesehen werden, welches von einer nichtlinearen Produktivitätszunahme ausgeht und neun Produktivitätsfaktoren identifiziert (s. Abbildung 4): Schülermerkmale (Fähigkeit, Entwicklungsstand und Motivation), Merkmale des Unterrichts (Unterrichtsqualität und -quantität) sowie Umweltmerkmale (familiäre Bedingungen, Klassenklima, Gruppe der Gleichaltrigen und Massenmedien). Die Annahme der nichtlinearen Produktivitätssteigerung zieht ein begrenztes Kompensationspotential der einzelnen Faktoren mit sich, d. h. Defizite lassen sich nur begrenzt durch Produktivitätsverbesserungen anderer Faktoren ausgleichen. Die ersten fünf Produktivitätsfaktoren sind dabei als absolut notwendige Schulleistungsvoraussetzungen nicht gegenseitig kompensierbar und gleichsam die Kernelemente der Modelle von Carroll und Bloom.

Allen bisher vorgestellten Modellen gemein ist, dass die institutionelle Einbindung der Unterrichtsprozesse fehlt und kontextuelle Variablen nicht in das Modell einfließen. Walberg (1981) greift diesen Aspekt im Sinne der Einbettung der Schulklasse als soziales System in das übergeordnete System Schule auf. Die zeitlichen und curricularen Vorgaben aufgrund einer bestimmten Schulform und eines bestimmten regionalen sowie sozialen



Einzugsgebiets verdeutlichen zudem die Begrenztheit der Wirksamkeit von Lehrkräften (Fend, 1977). Walberg (1981) gibt diesbezüglich an, dass die prädiktive Validität hinsichtlich der Varianzaufklärung der Lernleistung durch die Unterrichtsqualität 15% beträgt und damit einen geringeren Einfluss hat als die kognitiven Fähigkeiten (60%), das Klassenklima (60%) und das familiäre Umfeld (40%). Aufgrund der Faktoreninterkorrelationen können die aufgeklärten Varianzanteile allerdings nur als grobe Indikatoren für den Faktoreinfluss interpretiert werden (Gruehn, 2000).

Blooms und Carrolls Modellen sowie deren Weiterentwicklungen gemein ist die Ausrichtung auf das Zielkriterium der Schulleistung. Walberg (1981) hingegen differenziert bereits auf zweiter Ebene in affektive, kognitive und behaviorale Lernziele. Die Schulleistung als einziges Qualitätskriterium schränkt das mögliche Zielspektrum schulischer Prozesse und in der Rückwirkung auch die Beschreibungsdimensionen von Unterrichtsqualität ein (Terhart, 2000). Im Fokus steht zwar die kognitive Entwicklung der Schüler, dabei sind gemäß Clausen (2002) die Förderung des Selbstkonzepts und des sozialen Lernens der Schüler sowie die Divergenzminderung innerhalb der Klasse ebenfalls Unterrichtsziele, die gleichzeitig auch ein Mittel zur Zielerreichung darstellen.

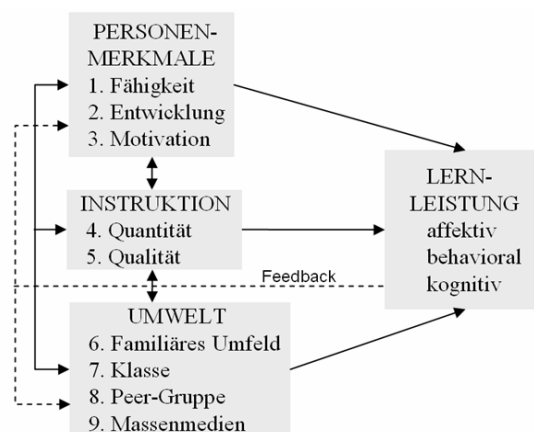


Abbildung 4: Produktivitätsmodell schulischen Lernens nach Walberg (in Anlehnung an Fraser, Walberg, Welch & Hattie, 1987, S. 158)

#### 2.2.4 Angebot-Nutzungs-Modell unterrichtlicher Wirkungen von Helmke

Das *Angebot-Nutzungs-Modell* von Helmke (2006a; 2007c, s. Abbildung 5) geht u. a. auf Fend (1981) zurück und integriert Merkmale der Unterrichtsqualität in ein umfassenderes Unterrichtsmodell, welches die Wirkungsweise und Zielkriterien des Unterrichts berücksichtigt. Die Grundannahme besteht darin, dass Unterricht als Angebot zu betrachten ist,

d. h. nicht unmittelbar den gewünschten Ertrag bzw. die intendierte Wirkung erzielt, sondern über die aktive Nutzung durch die Schüler mediert wird. Die Lernaktivitäten als Indikator der Unterrichtsnutzung resultieren dabei wiederum aus den Wahrnehmungs- und Interpretationsprozessen der Schüler hinsichtlich der Erwartungen der Lehrkraft und den Zielen unterrichtlicher Maßnahmen, mediert über emotionale, motivationale und volitionale Prozesse. Die Lernaktivitäten werden dabei, wie in den bisherigen Modellen schulischen Lernens bereits beschrieben, von den familiären Rahmenbedingungen und dem Lernpotential, welches die individuellen Eingangsvoraussetzungen umfasst, determiniert.

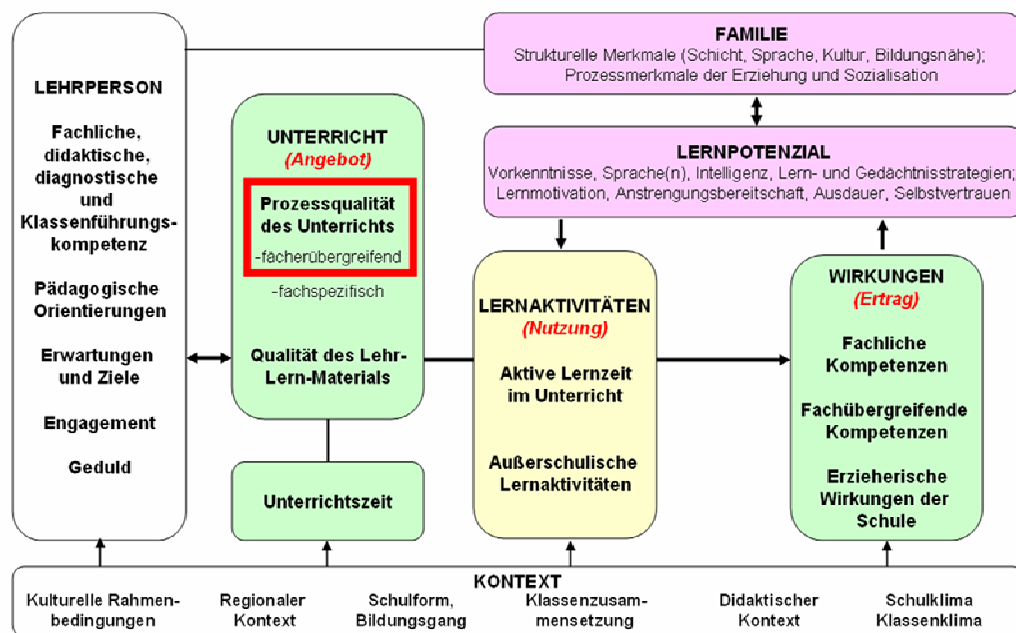


Abbildung 5: Angebots-Nutzungs-Modell von Helmke (2007c) mit Hervorhebung des im Fragebogen erfassten Modellausschnitts

Die unterrichtlichen Wirkprozesse sind insgesamt in die kontextuellen Gegebenheiten eingebettet. Entscheidende Determinanten des Unterrichtsangebots sind die Prozessqualität des Unterrichts und die Qualität des Lehr-Lernmaterials, wobei die Prozessqualität über fächerübergreifende sowie fachspezifische Merkmale mediert wird. Fächerübergreifende Prozessmerkmale sind: (1) Eine effiziente Klassenführung im Sinne eines unterrichtsbezogenen Zeitmanagements und der Etablierung von Regeln zur Vermeidung von Störungen und Zeitverlust, (2) ein lernförderliches Klima, welches den lernförderlichen Umgang mit Fehlern sowie die Herstellung einer entspannten und wertschätzenden Lernatmosphäre umfasst, (3) die thematische und instrumentelle Motivierung, z. B. durch das Aufzeigen der Relevanz des Lernstoffs für andere Unterrichtsfächer und den späteren Berufsalltag, (4) Klarheit und Verständlichkeit der Lehreräußerungen sowie die Strukturiertheit der Unter-

richtsstunde, z. B. durch das Vermitteln der Zielsetzungen zu Beginn oder Zusammenfassungen am Ende der Unterrichtsstunde, (5) Schülerorientierung im Sinne eines Ernstnehmens des Schülers und einer positiven Lernatmosphäre zur Förderung des Selbstvertrauens und der Lernbereitschaft, (6) Aktivierung, d. h. die Förderung des eigenständigen Lernens durch die Schüler, (7) Sicherung als Konsolidierung des Lernstoffs, (8) Wirkungsorientierung im Sinne des Einholens von unterrichtsbezogenem Feedback zur Standortbestimmung, (9) Passung in Form eines adäquaten Umgangs mit Heterogenität, z. B. durch die Variation fachlicher und überfachlicher Inhalte sowie die Anpassung des Tempos und des Schwierigkeitsgrads und (10) Methodenvielfalt, die in Abhängigkeit vom Fachkontext und den angestrebten Lernzielen realisiert werden sollte (genauer nachzulesen bei Helmke, 2006b; Helmke, 2007c).

Diese Merkmale werden von Helmke (2007c) als Orientierungsvariablen konzeptualisiert, bei denen die Wahrscheinlichkeit des Lernerfolgs mit der Anzahl an Variablen, die eine positive Ausprägung aufweisen, zunimmt. Die Unterrichtsmerkmale sind bedingt kompensierbar, d. h. sie sollten einen Grenzwert nicht unterschreiten, lassen sich aber durch hohe Ausprägungen anderer Merkmale ausgleichen.

Die Wirksamkeit der Prozessmerkmale des Unterrichts hinsichtlich der Entwicklung lern- und leistungsrelevanter Orientierungen, Einstellungen und dem Angstabbau sowie der Kompetenzentwicklung i. S. einer Qualitäts- und Quantitätserhöhung von Lern-, Gedächtnis- und Denkprozessen konnte in Studien gezeigt werden (Helmke, 2007c). Direkte Auswirkungen auf die Lernprozesse zeigen die Merkmale Klassenführung sowie die Klarheit, Strukturiertheit und Verständlichkeit des Unterrichts (Helmke & Schrader, 1993), dabei sind Klassenführung und Strukturiertheit besonders wichtig. So wies in der Studie keine Klasse mit einer hohen Leistungssteigerung unterdurchschnittliche Werte bei diesen Merkmalen auf. Die längsschnittlich angelegte Münchner Hauptschulstudie zeigte einen direkten Effekt von  $r = .19$  vom Unterrichtsmerkmal Klarheit und von der Zeitnutzung einen von  $r = .29$  auf die Mathematikleistung am Ende der 5. Klasse (vgl. Helmke & Weinert, 1997, S. 145). Der Einfluss der Klassenführung in Form effektiver Zeitnutzung und Störungsprävention ist bisher am besten hinsichtlich der Wirksamkeit auf die schulische Leistung belegt (Brophy & Good, 1986; Creemers, 1994; Wang, Haertel & Walberg, 1990). Clausen, Reusser und Klieme (2003) weisen darauf hin, dass dieser Befund auch darauf zurückzuführen sein könnte, dass die anderen Merkmalsbereiche im Gegensatz zur Klas-

senführung bzw. der Instruktionseffizienz weniger gut im konkret beobachtbaren Verhalten verankert und damit schwerer zu beurteilen sind. Die Unterrichtsmerkmale lernförderliches Klima, Schülerorientierung und -motivierung weisen indirekte Auswirkungen auf kognitive Zielkriterien über motivationale Zustände mediiert auf. Wie eingangs erwähnt, zeigt Gruehn (1995) in einer Studie, dass Disziplin, Klarheit des Unterrichts und eine individuelle Bezugsnormorientierung als Unterrichtsmerkmale sowohl mit der Leistungsentwicklung als auch mit emotional-motivationalen Zielkriterien, über das Selbstkonzept und die Schulfreude operationalisiert, positiv korrelieren.

Das Modell vereint eine Reihe der Befunde zur Unterrichtsforschung und unterstreicht, dass Schulleistungen das Resultat der Interaktion vieler Faktoren sind. Ebenso schärft es den Blick für die Bidirektionalität des Angebots-Nutzungs-Pfads, d. h. auch der Schülerinput kann als Lernangebot für die Lehrkraft verstanden werden, denn Schülerfehler können z. B. zum Überdenken der Lehrinhalte und -methoden anregen (Helmke, 2007c). Ditton (2000) geht hinsichtlich der Produktebene noch einen Schritt weiter als Helmke. Er differenziert die Ergebnisse der schulischen und unterrichtlichen Wirkprozesse in die Bildungs- und Unterrichtsziele als Output sowie den beruflichen Erfolg und die gesellschaftliche Teilhabe als langfristige Wirkungen bzw. den sogenannten Outcome.

### **2.3 Strukturierungsansätze zu den Dimensionen der Unterrichtsqualität**

Aus den Modellen des Abschnitts 2.2 geht hervor, dass die Autoren unterschiedliche Unterrichtsmerkmale mit dem Konstrukt Unterrichtsqualität in Verbindung bringen (s. dazu auch die anglo-amerikanischen Sammelreferate, z. B. von Brophy & Good, 1986). Es mangelt somit in der Literatur an einer systematischen Dimensionierung des Konstrukts.

Nach Clausen, Schnabel und Schröder existierten 2002 mehr als 100 Konstrukte zur Erfassung unterrichtlichen Geschehens. Ditton (2002b) resümiert, dass „es eine Vielzahl an Kombinations- und Variationsmöglichkeiten von Methoden für guten und schlechten Unterricht gibt“ (S. 197). Dies zeigt sich auch in der Tatsache, dass die Definition von Unterrichtsqualität eine Funktion der folgenden miteinander in Wechselwirkung stehenden Bestimmungsgrößen ist: (1) Der gewählte Auflösungsgrad, d. h. ob die Mikro-, Meso- oder Makroebene bzw. die Input-, Prozess- vs. Produktebene fokussiert wird. Die Prozessebene fokussiert gemäß Helmke (2006a) die Lehr-Lern-Interaktionsprozesse im Unterricht und die Orientierung der Unterrichtsqualität an sachlichen Kriterien (z. B. Schülermotivierung).

Dies entspricht einer variablenorientierten Forschungsperspektive (Helmke, 2006a). Auf Produktebene hingegen stellen die gewählten kognitiven, motivationalen oder affektiven Zielkriterien den Bezugsmaßstab für das Qualitätsverständnis dar. Erfolgen die Identifikation erfolgreicher Lehrkräfte – gemessen an a priori festgelegten Zielkriterien – und die Analyse ihrer Verhaltensweisen bzw. der Wirkungsweise ihres Unterrichts, handelt es sich gemäß Helmke (2006a) um eine *personenzentrierte* Strategie zur Identifikation ‚guten‘ Unterrichts. (2) Dies führt zu einer weiteren Determinante der Definition von Unterrichtsqualität, denn Schul- und Unterrichtszielsetzungen sind multikriterial, d. h. kognitive vs. nicht-kognitive im Sinne motivationaler sowie affektiver Kriterien können genauso als Unterrichtsziele deklariert werden wie das Erreichen der Leistungsstärkeren vs. Leistungsschwächeren bzw. das Erzielen einer Divergenzminderung innerhalb der Klasse oder die Orientierung an einer sozialen, individuellen oder kriterialen Bezugsnorm. Das Verständnisniveau und die Lernatmosphäre werden z. B. bei einer individuellen Bezugsnormorientierung von Schülern höher eingeschätzt als bei einer sozialen (Krug & Lecybyl, 1999). (3) Die Definition der Unterrichtsqualität variiert in Abhängigkeit von der fokussierten Ziel- bzw. Schülergruppe aufgrund variierender Eingangsvoraussetzungen bzw. Merkmale der Schüler. Leistungsängstliche profitieren z. B. von direkten Instruktionsformen, selbstsichere und leistungsstärkere Schüler hingegen nicht unbedingt (vgl. Helmke, 2006a). (4) Die Lehrerpersönlichkeit spielt auch eine entscheidende Rolle, wobei die Lehrermerkmale wiederum mit Prozessmerkmalen interagieren. So zeigen Helmke, Schneider und Weinert (1986), dass eine hohe diagnostische Fähigkeit erst in der Interaktion mit einer hohen Strukturiertheit signifikant zur Leistungssteigerung der Schüler beiträgt. (5) Der fachspezifische vs. fächerübergreifende Fokus beeinflusst ebenfalls die Zusammensetzung des Kanons der Unterrichtsmerkmale. (6) Neben der gewählten Perspektive auf die Unterrichtsqualität (Forscher, Politiker, Schulleitung, Kollegen, Lehrer, Schüler oder Eltern) spielt ferner (7) das zugrunde liegende normative Verständnis des Qualitätsbegriffs eine Rolle (Terhart, 2000).

Klieme und Rakoczy (2003) schlagen vor, die Vielfalt der Unterrichtsmerkmale durch eine Konzentration auf die Merkmale kognitive Aktivierung, Klassenführung bzw. Zeitnutzung und Schülerorientierung zu reduzieren. Einsiedler (1997, S. 228) rät zur systematischen Aufschlüsselung des Instruktionsverhaltens in Makromethoden des Unterrichts (z. B. entdeckungsorientiertes Vorgehen, dozierendes Vorgehen), Mikroverhaltensweisen der Lernsteuerung (z. B. Strukturierungshinweise, Feedback), Sozialformen (Klassenunterricht, Einzelarbeit, Gruppenarbeit) und Sozialklima. Wang, Haertel und Walberg (1990; 1993),

die auf der Grundlage von Metaanalysen und Expertenbefragungen eine Strukturierung vornehmen, weisen eine recht hohe inhaltliche Übereinstimmung mit Einsiedler auf. Gemäß einer Gegenüberstellung von Clausen et al. (2002, S. 248) können allerdings u. a. die Makromethoden bei Wang et al. mit den Aspekten Unterrichtsquantität, der Schaffung und Kontrolle der Rahmenbedingungen bezüglich der Lernzielumsetzung und durchzuführenden Leistungs- und Lernzielkontrollen stärker differenziert werden. Clausen et al. (2002) nähern sich einer systematischen Ordnung der Konstrukte über die Verfahren einer multidimensionalen Skalierung und Clusteranalyse im Anschluss an die expertenbasierte Zuordnung der Konstrukte zu zusammengehörigen Gruppen, wobei auch die Konstrukte der Unterrichtsklimaverfahren (s. Abschnitt 3.4) berücksichtigt werden. Diese Vorgehensweise ergibt acht Konstruktgruppen, die in Tabelle 1 den Merkmalen guten Unterrichts nach Helmke (2006a) gegenübergestellt sind.

Tabelle 1: Gegenüberstellung der von Clausen et al. (2002) ermittelten Clustergruppen sowie beispielhafter Indikatoren zu den von Helmke (2006b) postulierten 10 Merkmalen guten Unterrichts, wobei die im Fragebogen erfassten Merkmale hervorgehoben sind

Clustergruppen (Clausen et al., 2002, S. 256-257)	Beispielhafte Konstruktindikatoren für die jeweilige Clustergruppe	Merkmale guten Unterrichts (Helmke, 2006b, s. Abschnitt 2.2.4)
(1) Sozialorientierung des Unterrichts bzw. des Lehrers	Wertschätzung und Unterstützung der Schüler durch die Lehrkraft	(5) Schülerorientierung, (2) <b>lernförderliches Klima</b>
(2) Sozialklima	positive vs. negative Charakteristika der Schülersozialbeziehungen	
(3) Einstellungen und Lernhaltungen der Schüler	Leistungsbereitschaft, Desinteresse sowie allgemeine Unterrichtszufriedenheit	
(4) didaktische Qualität des Unterrichts	Klarheit und Strukturierung sowie kognitive Aktivierung	(4) <b>Klarheit und Verständlichkeit</b> , (10) Methodenvielfalt
(5) Effizienz der Klassenführung	effiziente und ineffiziente Aspekte der Klassenführung	(1) <b>effiziente Klassenführung</b> , (7) Sicherung
(6) Zeitnutzung im Unterricht	Aufgabenorientierung und Unterrichtstempo vs. Leistungsüberforderung	
(7) Motivationsunterstützung im Unterricht	Differenzierung und Förderungsorientierung	(3) <b>Motivierung</b> , (9) Passung/ Umgang mit Heterogenität
(8) Lehrerzentriertheit bzw. Schülermitbestimmung	Lehrerlenkung, Kontrolle vs. Schülerbeteiligung	(6) <b>Aktivierung</b>
		(8) Wirkungsorientierung

Es zeigt sich eine recht hohe Übereinstimmung zwischen den Konstruktgruppen, wobei Helmke (2002) bezüglich des Vorgehens von Clausen et al. (2002) zum Ausschluss des Klassen- bzw. Sozialklimas und Facetten der Mitarbeit sowie Lernbereitschaft als Folgen und Bedingungen des Unterrichts rät, da sie keine Unterrichtskonstrukte im engeren Sinn darstellen. Nicht in den Konstrukten von Clausen et al. (2002) aufgeführt, sind Aspekte der Wirkungsorientierung im Sinne einer systematischen feedbackbasierten Standortbestimmung durch die Lehrkraft. Werden diese Strukturierungen auf die im Kooperationsprojekt vereinbarten zu erfassenden Fragebogenkonstrukte übertragen, zeigt sich eine hohe Überlappung (vgl. Tabelle 1). Als Teilmenge der von Clausen et al. (2002) ermittelten Clustergruppen erfassen die im Fragebogen verwendeten Unterrichtsmerkmale hinsichtlich der Skalierungsdimension Sozial- vs. Leistungsorientierung den Pol der effektivitäts- bzw. leistungsorientierten Instruktion. Bezüglich der Dimension Subjekt-Objekt werden v. a. die (Schüler)<-Lehrer- sowie (Lehrer)<-Lehrer-Richtungen abgedeckt. Somit werden Merkmale erfasst, deren Objekt im Wesentlichen der Schüler ist, z. B. bei Motivierungs- und Aktivierungsinterventionen und Merkmale, wie z. B. die Klassenführung bzw. Klarheit und Verständlichkeit, die auf den Lehrer selbst bzw. den Unterricht abzielen. Im nächsten Kapitel folgen Überlegungen zur Unterrichtswahrnehmung aus der Schülerperspektive sowie ihrer Operationalisierung.

### 3 Unterrichtswahrnehmung aus Schülersicht

Im Gegensatz zur frühen Lehr-Lernforschung, die eine Erfassung der Merkmale effektiven und erfolgreichen Unterrichts im Wesentlichen auf der Basis objektiver Beobachtungsdaten anstrebt, rückt die Klimaforschung die Schülerwahrnehmung als konstituierendes Merkmal des Unterrichts- und Schulklimas in das Zentrum ihrer Forschungstätigkeit. Im nächsten Abschnitt werden deshalb wesentliche Ansätze der Klimaforschung vorgestellt (Abschnitt 3.1). Danach wird die Schülerwahrnehmung als interpersonaler Prozess betrachtet (Abschnitt 3.2). Im Abschnitt 3.3 werden Vor- und Nachteile der Schülerperspektive als Datenquelle thematisiert, bevor das Kapitel mit Überlegungen zur Operationalisierung der Schülerwahrnehmung schließt (Abschnitt 3.4).

#### 3.1 Schülerwahrnehmung als Fokus der Klimaforschung

Der Unterrichtseinfluss lässt sich aus der Sicht von Klimaforschern nicht allein anhand organisierter Lernprozesse beschreiben, denn Unterricht stellt aus der Klimaforschungsperspektive eine soziale Umwelt mit unterschiedlichen „klimatischen Bedingungen“ dar (vgl. Eder, 1996; Fend, 1977; Lange et al., 1983; Moos, 1979). Nach der Darstellung der Grundlagen der Klimaforschung (Abschnitt 3.1.1) wird die Schülerwahrnehmung als eine dieser klimatischen Bedingungen betrachtet (Abschnitt 3.1.2).

##### 3.1.1 Wurzeln der Klimaforschung

Wesentliche Anfangsimpulse zur Klimaforschung stammen aus *Lewins Feldtheorie* und *Lebensraumkonzept* (1936) sowie *Murrays interaktionspsychologischem ‚Need-Press-Modell‘* (1938, zitiert nach Gruehn, 2000), die hier in Anbetracht einschlägiger Übersichten (Dreesmann, 1978; Eder, 1996) nur skizziert werden sollen.

*Lewins* (1936) Ansätze charakterisieren (1) das Individuum als psychologische Person, (2) die für die Person bedeutsamen Weltausschnitte als psychologische Umwelt und (3) das Wissen über diese Umwelt als drei Merkmale, deren Verschränkung das psychologische Feld konstituiert. Person und Umwelt bedingen sich somit gegenseitig und dieselbe Umwelt kann von mehreren Individuen unterschiedlich wahrgenommen und gedeutet werden. Dass der Unterricht von den Schülern unterschiedlich erlebt werden kann, prägt die Klimaforschung dahingehend, dass die subjektive Wahrnehmung situativer Bedingungen, vermittelt über motivationale, soziale und wertorientierte Unterrichts- und Lehrermerkmale als Prä-



diktor für das Verhalten in den Fokus rückt. Der Brechung und Spiegelung der Umweltbedingungen in den Schülerkognitionen und dem -erleben wird somit pädagogische Relevanz attribuiert (vgl. Dreesmann, 1982). Die Wahrnehmung der vorhandenen Umwelteinflüsse werden bei Murray (1938, zitiert nach Gruehn, 2000) als *beta press* bezeichnet, die zusammen mit den *alpha press*, d. h. den gegebenen Umwelteinflüssen und den individuellen Wünschen bzw. Bedürfnissen (needs) das Verhalten des Individuums determinieren. Stern, Stein und Bloom (1956) differenzieren *beta press* weiter in *private beta press* als idiosynkratische Umweltwahrnehmung und *consensual beta press*, als von Gruppenmitgliedern aufgrund gemeinsamer Denkweisen und Erfahrungen geteilte Umweltwahrnehmung.

Den Ausgangspunkt für die Erforschung des Klimas in Schulklassen bildet das *Modell sozialisatorischer Einflüsse* von Getzels und Thelen (1960). Die Gruppe fungiert dabei, wie in Abbildung 6 ersichtlich, als „soziales Bindeglied und Vermittlungsinstanz zwischen Individuum und Gesellschaft“ (Eder, 1996, S. 36) und mediiert dadurch zwischen der sozialen Rolle innerhalb einer Institution auf der nomothetischen Dimension und der Persönlichkeit auf der idiographischen Dimension. Eine Reihe von Klimaforschern setzen bei der Erfassung des Gruppenklimas als Untersuchungseinheit an, weil es die individuellen und strukturellen Gegebenheiten in sich verschränkt und die Vorhersage sozialen Verhaltens erlaubt.

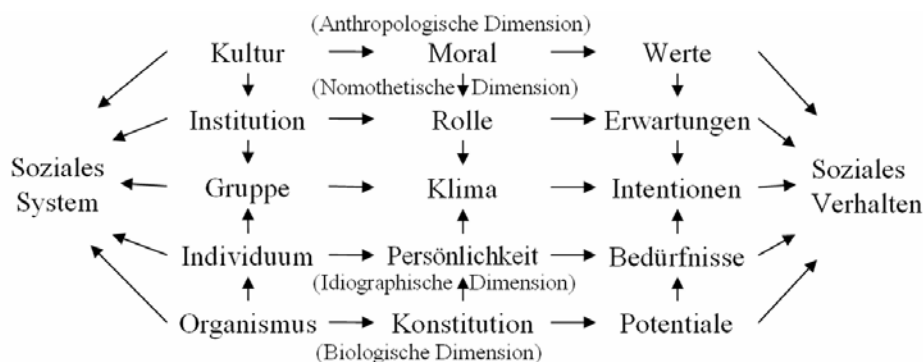


Abbildung 6: Erweitertes Modell sozialisatorischer Einflüsse (Getzels & Thelen, 1960)

Auch der *sozialisierungstheoretische Ansatz* von Fend (1977) greift diesen Aspekt in seiner Definition des Schulklimas als „die Art und Weise, wie Sozialisationsprozesse in veranstalteter Form durchgeführt werden, die „Verlebendigung“ institutioneller Verhältnisse durch die Individualität der Lehrer und Schüler und die dabei entstehenden Lebensformen“ (S. 64) wieder auf und thematisiert den Einfluss kollektiver Erfahrungen und struktureller Merkmale in der Klasse auf die Entstehung einer sozialen Persönlichkeit sowie eines damit einhergehenden sozialen Verhaltens.

### 3.1.2 Schülerwahrnehmung als eine Konstituente des Klimabegriffs

Zur Zeit dominiert das Verständnis des Klimabegriffs gemäß Eder (2001) in seiner Verwendung zum Erfassen und Beschreiben subjektiv wahrgenommener Lernumwelten. Dabei weist der Klimabegriff zwei konstituierende Definitionsmerkmale auf: (1) Untersuchungsgegenstand sind die Wahrnehmungen wichtiger Organisationsmerkmale sowie die auf dieser Basis gebildeten Überzeugungen und (2) diese werden anhand von Selbstberichten erfasst (Eder, 2001). Auf den schulischen Kontext übertragen, bedeutet Klima somit das „Insgesamt schulischer Merkmale in der Wahrnehmung der Schüler“ (Lange et al., 1983, S. 11) bzw. „die (subjektive) Wahrnehmung, die (subjektive) Beurteilung, das (subjektive) Erleben schul- und unterrichtsbezogener Aspekte durch den Schüler“ (Arbinger & von Saldern, 1984, S. 88). Dabei ist der Klimabegriff jedoch nicht auf eine subjektive Perspektive der Schüler beschränkt, bestimmte Unterrichtsmerkmalskonfigurationen können mit Bezug auf den Klimabegriff auch von den Lehrkräften bzw. unbeteiligten Beobachtern eingeschätzt werden. Eder (1996; 2001) rät zur Differenzierung des Klimabegriffs anhand der fünf Facetten Inhalt, Organisationsbezug, Subjektbezug, Aggregierungsebene und Quelle (s. Tabelle 2).

Tabelle 2: Facetten zur Eingrenzung des Klimabegriffs (Eder, 1996; 2001)

Inhalt	Organisationsbezug	Subjektbezug	Aggregierungsebene	Quelle
Sozialklima	Schulklima	<b>Psychologisches Klima</b>	Individualklima	<b>Schülerklima</b>
<b>Unterrichtsklima</b>	Lehrkörperklima	<b>Kollektives Klima</b>	<b>Aggregiertes Klima</b>	<b>Lehrerklima</b>
Werteklima	<b>Klassenklima</b>			Elternklima
u.ä.	Gruppenklima			
	u.ä.			

*Anmerkung:* Die Hervorhebungen kennzeichnen die Charakteristiken des in der Arbeit verwendeten Fragebogenverfahrens.

Der *Inhaltsaspekt* nimmt dabei auf die erfassten Sachverhalte Bezug, so können z. B. Schülerbeziehungen, Werte oder Unterrichtsmerkmale in den Klimabeschreibungen erfasst werden. Der *organisationale Bezug* spiegelt die Tatsache wider, dass sich Klimawahrnehmungen auf die gesamte Organisation oder Ausschnitte dieser beziehen können. Der *Subjektbezug* thematisiert, dass Klimawahrnehmungen die subjektiv erlebte Umwelt (psychologisches Klima) einer einzelnen Person oder gemeinsame Wahrnehmungen einer beurteilenden Personengruppe erfassen können (kollektives Klima). Die *Aggregierungsebene* unterscheidet das individuelle Klima, welches Aufschluss über die Wahrnehmung des

Klimas durch eine Person gibt von dem aggregierten Klima, das den durchschnittlichen Wahrnehmungswert einer Personengruppe repräsentiert. Welche Personengruppe die Klimawahrnehmungen liefert, wird in der Definition von Eder (1996; 2001) durch das Merkmal *Quelle* gekennzeichnet.

Zu den von der Klimaforschung erfassten Klimadimensionen zählen auch Unterrichtsmerkmale, die mit dem zu überprüfenden Fragebogen erhoben werden (s. dazu Abschnitt 3.4). Wird die ‚Klimawahrnehmung‘ in diesem Sinne verstanden, lässt sich das Fragebogenverfahren hinsichtlich der Terminologie von Eder (1996; 2001) folgendermaßen einordnen: Es dient inhaltlich der Erfassung des ‚Unterrichtsklimas‘ als das aufgrund der Interaktion der Lehrkraft eines bestimmten Fachs und ihres jeweiligen Unterrichtsstils mit den Schülern bzw. ihren individuellen Verhaltensweisen und Lernstilen hervorgerufene Klima. Die Klimawahrnehmungen beziehen sich dabei auf die Organisationseinheit Klasse und repräsentieren sowohl das psychologische als auch das kollektive Schülerklima anhand des Schülerfragebogens sowie das psychologische Lehrerklima mit Hilfe des Lehrerfragebogens operationalisiert, welche anschließend in einer aggregierten Form rückgemeldet werden.

Wie aus den individuellen Unterrichtswahrnehmungen ein kollektives (Klassen-)Klima entsteht, beleuchtet Dreesmann (1979; 1982) mit seinem *sozial-kognitiven Modell*, welches allgemeine Kommunikationsprozesse mit den Postulaten des dynamischen Interaktionismus verbindet. Dreesmann postuliert, dass zwei Prozesse an der Entstehung eines Klassenklimas beteiligt sind: (1) Die Verarbeitung des Unterrichts auf kognitiver Ebene durch die einzelnen Schüler und (2) ein gruppendynamischer Prozess, der über gemeinsame Erlebnisse, Interaktionen und Abgrenzungsverhalten gegenüber anderen Klassen zum Aufbau gemeinsamer Erlebens- und Verarbeitungsstrukturen führt. In den kognitiven Verarbeitungsprozess, welcher zu affektiven, attributiven und erwartungsbezogenen Unterrichtsurteilen führt, fließen Personenmerkmale und aus der Unterrichtsinteraktion resultierende situative Merkmale ein. Die situativen Merkmale nehmen hinsichtlich des Schülerwahrnehmungsprozesses ein höheres Gewicht ein, woraus sich ergibt, dass die Übereinstimmung zwischen Schülern innerhalb einer Klasse größer ist als zwischen Klassen. „Diese Umwelt wird von jedem individuell erlebt, doch hat dieses Erleben einen Teil, der allen Schülern in der Klasse gemeinsam ist...“ (Dreesmann, 1979, S. 51). Der gemeinsame Erlebensanteil entspricht dem Unterrichtsklima, quasi im Sinne des *consensual beta press* nach Stern et al.

(1956). Aus dieser Definition ergibt sich, dass zur Erfassung der Klimas der Klassenmittelwert allein nicht ausreicht, da die individuellen Wahrnehmungen Konstituenten des kollektiven Klimas sind. In der Konsequenz muss die Streuung der Unterrichtswahrnehmung ebenfalls berücksichtigt werden, die in ihrer extremen Ausprägung ein Hinweis darauf wäre, dass der kollektiv geteilte Erlebnisanteil gering ist (vgl. Gruehn, 2000). Dies entspricht den Ausführungen von Clausen (2002), nach dem der Klassenmittelwert nur bedingt als reliabler Indikator der Schülerwahrnehmungen angesehen werden kann, denn in der ‚Fehlervarianz‘ stecken nicht nur Messfehler im klassischen Sinn, sondern auch spezifische Varianzanteile, die bei der Aggregation an Gewicht verlieren.

Im Folgenden wird von Unterrichtswahrnehmung gesprochen, da Clausen (2002) darauf hinweist, dass der Begriff Unterrichtsklima einen stark metaphorischen Charakter aufweist, genau wie der Begriff Unterrichtsqualität die Subjektivität der zugrunde liegenden Datenquellen vernachlässigt. Gemäß Dreesmann (1979) sind Unterrichtswahrnehmungen „kognitive Prozesse als situativ beeinflusste personale Merkmale“ (S. 80), daher soll im nächsten Abschnitt aus der Perspektive der interpersonalen Wahrnehmungsforschung auf die Determinanten und Wirkprozesse von Schülerwahrnehmungen geschaut werden.

### **3.2 Schülerwahrnehmung als interpersonaler Wahrnehmungsprozess**

Unterrichtswahrnehmungen sind *subjektiv*, denn sie sind – einer allgemeinen Definition von Wahrnehmung folgend – das Resultat einer selektiven Aufnahme von Umweltreizen sowie deren emotionale und kognitive Verarbeitung in Abhängigkeit von individuellen Bedürfnissen, Wünschen, Zielen und Fähigkeiten des Wahrnehmenden. Als *soziale* Wahrnehmungsprozesse (Clausen, 2002) werden sie auch durch die Ziele, Bedürfnisse, Einstellungen und Wahrnehmungen der Interaktionspartner beeinflusst. Eine nähere Aufschlüsselung der individuellen und sozialen Determinanten der Unterrichtswahrnehmung erlaubt das *PERSON-Modell zur interpersonalen Wahrnehmung* von Kenny (2004). Hinsichtlich der Übertragung der Modellimplikationen auf den Schulkontext ist kritisch anzumerken, dass die durch das Modell integrierten Studienergebnisse im Wesentlichen auf der Einschätzung von Persönlichkeitsmerkmalen (z. B. Big Five) beruhen und von daher nicht ungeprüft auf die Einschätzung der Unterrichtsqualität übertragen werden können. Das Modell weist allerdings für die bei der Unterrichtsbeurteilung durch den Schüler ablaufenden Wahrnehmungs- und Urteilsprozesse einen hohen heuristischen Wert auf.

Kenny nimmt zwei auf kategorialer Information beruhende und vier verhaltensbasierte Varianzquellen einer Personenbeurteilung an: *Stereotype* bilden geteilte Wahrnehmungen ab, die auf einer kategorialen Information, z. B. der physischen Erscheinung beruhen (z. B. Geschlecht der Lehrkraft), wohingegen sich *Residual* auf den idiosynkratischen und somit nicht-geteilten Wahrnehmungsanteil auf der Basis persönlicher Stereotype bezieht (z. B. Schüler x schreibt männlichen Lehrkräften einen offeneren Umgang mit Schülervorschlägen zu). Die Variablen *Personality* als das von den Beurteilern aufgrund einer repräsentativen Verhaltensstichprobe gewonnene Persönlichkeitsurteil (z. B. eine Lehrkraft geht sehr offen mit neuen Ideen und Vorschlägen der Schüler um) und *Norm* erfassen ebenfalls geteilte Wahrnehmungen. Die Variable *Norm* bildet das vom allgemeinen Persönlichkeitsurteil abweichende Verhalten der zu beurteilenden Person in ihrer unique Bedeutung ab, die die Beurteiler einem singulären Verhaltensakt übereinstimmend zuschreiben (obwohl die Lehrkraft z. B. im Allgemeinen als offen eingestuft wird, kann sie in einer bestimmten Situation von der Klasse übereinstimmend als konventionell bzw. konservativ hinsichtlich der Schülervorschläge beurteilt werden). Wagner (2008) weist darauf hin, dass die Betrachtung der Norm-Komponente als vollkommen geteilte Wahrnehmung nur dann gerechtfertigt ist, wenn die zugrunde liegende Verhaltensstichprobe allen Beurteilern gleichermaßen zur Verfügung steht. Diese Bedingung wäre bei Schülern, die erst im Laufe des Schuljahres in die Klasse kommen, krankheitsbedingt eine gewisse Zeit fehlen oder Sequenzen innerhalb der Unterrichtsstunden verpassen, nicht erfüllt (vgl. Wagner, 2008). Nicht-geteilte Wahrnehmungen finden sich ferner in der *Opinion*-Variable als individuelle verhaltensbasierte allgemeine Wahrnehmung der zu beurteilenden Person (Schüler x stuft das Lehrkraftverhalten als offen ein, wohingegen Schüler y dieses weniger offen einschätzt) und der Variable *Error*, die den Wahrnehmungsanteil charakterisiert, der von der allgemeinen Verhaltenswahrnehmung abweicht, aber im Gegensatz zur Variable *Norm* idiosynkratisch ist (Schüler x sieht – im Gegensatz zu den anderen Schülern – die Information, dass die Lehrkraft vernarrt in Tiere ist, als einen Indikator für ein potentiell antisoziales Verhalten). Kennys Modellierungen zeigen, dass die anfänglichen Beurteilungsprozesse von den auf kategorialen Informationen beruhenden Stereotyp- und Residual-Komponenten dominiert werden, wohingegen deren Varianzanteil mit zunehmender Anzahl beobachteter Verhaltenssequenzen abnimmt. Bereits nach ca. 90 beobachteten Verhaltenssequenzen bzw. nach maximal einem halben Schuljahr – wenn eine Verhaltenssequenz konservativ mit einer Unterrichtsstunde gleichgesetzt wird – ist die Personenwahrnehmung im Wesentlichen eine

Funktion der Personality- und Opinion-Komponente (vgl. Kenny, 2004, S. 271). Die Schüler rekurren in ihren Unterrichtswahrnehmungen somit auf die relativ stabilen Trait-Komponenten auf der Ebene beobachtbaren Verhaltens der Lehrkraft. Die Befragung der Schüler sollte hinsichtlich des *Zeitpunkts* somit dann durchgeführt werden, wenn die Unterrichtswahrnehmung im Wesentlichen durch die Opinion- (auf der Individualebene) und die Personality-Komponente (auf Klassenebene) bestimmt wird, d. h. wenn die Schüler bereits eine größere Anzahl an Verhaltenssequenzen der Lehrkräfte beobachten konnten.

Ferner zeigt Kenny (2004, S. 275), dass die geteilte Personality-Wahrnehmungskomponente mit zunehmender Anzahl von Beurteilern (d. h. Schülern) bei einer *Aggregation der Daten* die nicht-geteilte Opinion-Komponente dominiert. Daher sollte das Erfassen möglichst aller Schüler einer Klasse und die Datenaggregation angestrebt werden. Die Anwendung der Aggregation trägt auch der Forderung nach Anonymität der Schülerangaben und einer Reduktion der Ergebniskomplexität im Kontext der Unterrichtsentwicklung Rechnung. Kenny (2004) gibt zu bedenken, dass der Effekt der Aggregation durch die Kommunikation der Beurteiler reduziert wird – dies ist im schulischen Kontext eines Klassenverbands sehr wahrscheinlich und vor dem Hintergrund überfachlicher Entwicklungskriterien erwünscht. Die Kommunikation könnte eine nicht notwendigerweise mit dem realen Bild korrelierende Konstruktion einer geteilten Personeneinschätzung bewirken (Wagner, 2008). Der dabei steigende Anteil der Stereotyp-Komponente könnte eventuell durch einen höheren Bekanntheitsgrad der zu beurteilenden Person ausgeglichen werden. Die Voraussetzungen im Kontext der Schülerbefragungen sind dafür sehr gut. Welche weiteren Vorteile die Unterrichtseinschätzung aus der Schülerperspektive aufweist, wird im nächsten Abschnitt betrachtet.

### 3.3 Schüler als Datenquelle

Die Unterrichtsqualität lässt sich lediglich indirekt erfassen (Clausen, 2002). Eine solche Einschätzung kann sowohl von den Unterrichtsteilnehmern, d. h. der Lehrkraft und den Schülern selbst als auch durch Videoaufzeichnungen und außenstehenden Beobachtern z. B. durch externe oder interne Unterrichtsbeobachter im Rahmen von Unterrichtshospitationen bzw. -besuchen vorgenommen werden. Das Erfassen der Unterrichtsqualität aus der Schülerperspektive weist einige *Vorteile* auf: Im Vergleich zur Unterrichtbeobachtung durch Dritte sind Schülerangaben eine sehr effiziente und ökonomische Datenquelle, denn sie erlauben mit relativ geringem Aufwand und somit kostengünstig, einen Datenpool im

Umfang der jeweiligen Klassengröße zeitgleich zu erheben (vgl. Clausen, 2002; De Jong & Westerhof, 2001; Helmke, 2006a). Bei externen Unterrichtsbeobachtungen muss hingegen aus Reliabilitätsgründen versucht werden, zumindest zwei oder mehr Beobachter einzusetzen, was mit einem erheblichen organisatorischen und logistischen Aufwand einhergeht. Gleichsam fungiert die Beobachteranwesenheit als Störfaktor und erhöht die Fehlervarianz, wohingegen dies bei einer Schülerbefragung nicht der Fall ist (Steltmann, 1992). Auch die Subjektivität externer Beobachter charakterisiert Steltmann (1992) als Verzerrung der Messgenauigkeit, denn diese sei mit dem Unterricht unkorreliert, wohingegen die subjektive Wahrnehmung des Lehrerverhaltens aus Sicht der Schüler als pädagogisches Feldelement das Unterrichtsklima prägt. „Nicht ob ein Lehrer tatsächlich „freundlich“ ist (nach Maßstäben, die von außen angelegt werden), ist entscheidend für das Verhältnis zwischen Lehrer und Klasse, sondern ob die Klasse seine Art der Zuwendung als freundlich empfindet“ (Steltmann, 1992, S. 571). Die Schülerwahrnehmung fungiert somit im pädagogischen Prozess als mediiender Faktor zwischen den Verhaltensweisen des Lehrers als Inputvariable und den Schülerverhaltensweisen sowie -lernleistungen als Outputvariablen. Helmke, Schneider und Weinert (1986) modellieren entsprechend in ihrem Vorhersagemodell der Schülerleistung die Schülerwahrnehmung als kognitive Mediatorvariable zwischen beobachteten Unterrichtsmerkmalen und den Lernleistungen der Schüler.

Die spezifischen Eigenschaften des Unterrichts als Beurteilungsgegenstand mediiieren weitere Vorteile der Schülerperspektive, denn Unterricht wird von Clausen (2002) als „eine spezifische Lernumwelt, eine sich wiederholende Situation, die durch die Interaktion von Lehrer- und Schülerverhalten gestaltet wird und typische, zeitlich überdauernde Qualitäten aufweist“ (S. 81) verstanden. Dementsprechend beruhen die Schülerangaben auf einem relativ langen Beobachtungszeitraum und somit „...einer längeren Phase der Vertrautheit...mit der realen Unterrichtssituation“ (Helmke, 2007a, S. 99). Schüler können somit auf ein differenziertes Beobachtungsspektrum zurückgreifen, denn sie kennen ihre Lehrkräfte sowohl im interindividuellen Vergleich mit anderen Lehrkräften als auch ihre intraindividuell variierenden Verhaltensweisen über verschiedene Situationen hinweg (Ditton, 2000). Da Schülerrückmeldungen im Gegensatz zu Fremdevaluationen nicht nur auf dem jeweilig beobachtbaren Verhaltensausschnitt basieren, werden potentielle Verhaltensänderungen oder -akzentuierungen während des zu evaluierenden Unterrichtsausschnitts in den Unterrichtseinschätzungen tendenziell abgeschwächt.

Aufgrund der Tatsache, dass das (Urteils-)Verhalten nicht auf objektiven Stimuli sondern auf der subjektiven Repräsentation dieser beruht (Bless, Fiedler & Strack, 2004), unterliegen allerdings auch Schüler typischen psychologischen *Beurteilungsprinzipien und -phänomenen*, die den Prozess von der Wahrnehmung bis hin zum Verhalten, z. B. durch eine selektive Wahrnehmung, Informationsverarbeitung und -abruf aufgrund einer begrenzten Verarbeitungskapazität mediiieren. Schülerangaben können bei der Formatierung ihrer Antworten auch *potentiell verzerrende Antworttendenzen* aufweisen (vgl. Dreesmann, 1982; Mummendey, 1995): Die Akquieszenz, d. h. die Ja-Sage- bzw. Zustimmungstendenz kennzeichnet die Neigung der beurteilenden Personen, dem Iteminhalt eher zuzustimmen, als diesen abzulehnen. Die Präferenz extremer Skalenwerte beschreibt die Extremtendenz, die sich im Rahmen der Schülerbefragung z. B. in Milde- oder Strengeneffekten ausdrücken kann. Die Mittetendenz bzw. zentrale Tendenz stellt das Gegenteil der Extremtendenz dar und bezeichnet die Neigung, mittlere Antwortvorgabe zu wählen. Die Neigung zu sozial erwünschtem Antwortverhalten bedeutet, dass die beurteilende Person entgegen der eigenen Auffassung entsprechend der sozialen Norm oder den Erwartungen anderer Personen, z. B. der Lehrkraft oder der Peer-Gruppe innerhalb des Klassenverbands, antwortet. Schließlich kann bei der Schülerbefragung auch der Halo-Effekt auftreten, der darin besteht, faktisch unabhängige oder mäßig korrelierte Aspekte von Beurteilungsgegenständen als zusammenhängend bzw. nicht in ausreichendem Maße differenziert wahrzunehmen (Thorndike, 1920). Eine verbreitete Auffassung zur Erklärung dieses Effekts ist, dass die Beurteilung eines salienten Einzelmerkmals die Beurteilung weiterer Merkmale überlagert. Gemäß Lance, LaPointe und Stewart (1994) passt die Erklärung, dass sich die urteilende Person einen summativen Gesamteindruck vom Beurteilungsgegenstand bildet, welcher sich in einer mit dem Gesamteindruck konsistenten Beurteilung der Einzelmerkmale widerspiegelt, am besten zu dem in ihrer Studie gefundenen empirischen Datenmuster.

Das Ausmaß des Verzerrungspotentials der Antworttendenzen sollte jedoch vor dem Hintergrund des jeweiligen Beurteilungskontextes reflektiert werden. Im Rahmen eines in die Schulkultur integrierten Schülerfeedbackansatzes zur Unterrichtsentwicklung wird die Schülerbefragung durch einen Ergebnisdialog ergänzt (s. Abschnitt 4.3). Der Austausch mit den Schülern ermöglicht die Reflexion der schriftlichen Schülerangaben auf verbaler Ebene anhand der noch zugänglichen und zur Urteilsfindung herangezogenen Verhaltenssequenzen. Clausen (2002) weist auch darauf hin, dass der Halo-Effekt eigentlich nur besagt, dass Schüler eine „...weniger differenzierte Wahrnehmungsstruktur aufweisen, als sie vom



Wissenschaftlicher intendiert oder in einer Eichstichprobe vorgefunden wurde“ (S. 45). Antwortverzerrungen kann ferner mittels eindeutig und verhaltensbezogen konstruierter Items und Antwortalternativen (Dreesmann, 1982, s. auch Abschnitt 4.5) sowie der Anonymitätswahrung bei der Durchführung (s. Abschnitt 4.3) entgegengewirkt werden. Steltmann (1992) resümiert daher, dass „...die Vorteile der Schülerbeurteilung die der Fremdbeurteilung doch bei weitem [überwiegen]. Die Möglichkeiten der Lehrerbeurteilung durch Schüler...allerdings in Deutschland noch kaum genutzt [werden]“ (S. 572).

In Bezug auf die Gültigkeit von Schülerurteilen unterstreichen Walberg und Haertel (1980) die prädiktive Validität von Schülerdaten zur Einschätzung der Lernumwelt hinsichtlich der Vorhersage von kognitiven, behavioralen und affektiven Lernergebnissen. Anderson und Walberg (1974) vergleichen dazu die Vorhersageleistung des LEI (Learning Environment Inventory, Fraser, Anderson & Walberg, 1982, zitiert nach Schröder, 1999) als Instrument zum Erfassen der Lernumwelt mit dem IQ als Prädiktor. 30% der Varianz kognitiver, affektiver und behavioraler Lernleistungen lassen sich durch die erfasste Schülerwahrnehmung des Klassenklimas erklären und nur 7% durch den Intelligenzquotienten der Schüler. In einer Synthese der Forschungsergebnisse zum Zusammenhang zwischen der Schülerwahrnehmung sozio-psychologischer Umweltmerkmale und den Lernergebnissen anhand der ermittelten Korrelationskoeffizienten kommen auch Hartel, Walberg und Haertel (1979, zitiert nach Walberg & Haertel, 1980) zu dem Fazit, dass Lernergebnisse, die sowohl die Schülerleistungen als auch Selbstkonzeptmerkmale umfassen, positiv mit der aus Schülersicht wahrgenommenen Klassenkohäsion, Aufgabenschwierigkeit, erlebten Demokratie und Zufriedenheit korrelieren. Negative Korrelationsmaße zeigen sich zum wahrgenommenen Konfliktpotential der Klasse, einer geringen Schülerbeteiligung, einer Tendenz zur Cliquenbildung und einer erlebten Desorganisation. Bürgisser (2007) gibt zu bedenken, dass die Übertragbarkeit des amerikanischen Schulsystems auf das deutsche noch offen ist, lediglich die College-Verhältnisse seien mit der Sekundarstufe II vergleichbar. Neuere Befunde unter Rückgriff auf die PISA-Daten zeigen allerdings auch signifikante Korrelationen zwischen der Schülereinschätzung der Klassenführung, des adäquaten Instruktionstempus, der wahrgenommenen kognitiven Autonomie und den mathematischen Schülerleistungen sowie der Zufriedenheit des Schülers mit dem Lehrer (Kunter & Baumert, 2006).

Die Schülerperzeptionen können durch verschiedene Personmerkmale auf der Individual-ebene beeinflusst werden, u. a. durch den Leistungsstand der Schüler (vgl. Marsh, 1984;

Moos, 1979). Bei guten Schulleistungen fallen die Schülereinschätzungen positiver aus als bei tendenziell schlechten Leistungen. Babad (1996) führt die differentielle Behandlung in Abhängigkeit vom Leistungsstand der Schüler als Erklärung dafür an. Dabei könnten auch Sympathie- oder Antipathietendenzen gegenüber der Lehrkraft einen vermittelnden Charakter einnehmen. Clausen (2002) weist darauf hin, dass sich entsprechende Effekte bei einer Aggregation auf Klassenniveau beim Vorliegen einer symmetrischen Leistungsverteilung innerhalb der Klasse ausmitteln. Gruehn (2000) zeigt in Übereinstimmung dazu, dass v. a. auf Klassenebene substantielle Zusammenhänge zwischen der Wahrnehmung der Lernumwelt und dem Leistungszuwachs bestehen, wohingegen die Zusammenhänge auf der Ebene des Schülers deutlich niedriger sind.

Zusammenfassend lässt sich festhalten, dass insbesondere der Schülersicht eine perspektivenspezifische Bedeutung beigemessen werden kann (Clausen, 2002), aber angesichts der aufgezeigten Grenzen Schülerfeedback nicht überwertet werden darf. Das Einbeziehen der Schülersicht hinsichtlich der Lernwirkung und Effektivität von Lehrmaßnahmen ist aber in jedem Fall angeraten, weil Schülereinschätzungen (1) „...Informationen über die Wirkungen und Nebenwirkungen von Lehrerverhaltensweisen..., die bei der Anwendung von Tests nicht (mit-)bedacht werden und die für die subjektive Sicht des Schülers, seine persönlichen Entwicklungen und Werthaltungen bedeutsam sind“ (Hofer, 1981, S. 50) liefern, (2) einfach und unmittelbar erhoben werden können und (3) eine Einsicht in die Erlebens- und Urteilsprozesse der Schüler als konstituierende Variable der bidirektionalen Lehrer-Schüler-Interaktion ermöglichen. Der Frage, wie Schülerwahrnehmungen operationalisiert werden können und welche spezifischen Messprobleme damit verbunden sind, widmet sich der nächste Abschnitt.

### **3.4 Operationalisierung von Schülerwahrnehmungen**

Aufgrund dessen, dass in der Klimaforschung die Wahrnehmung der Betroffenen ein konstituierendes Merkmal des Klimas ist, wurden in diesem Forschungsbereich eine Reihe von Verfahren zur Operationalisierung der Schülersicht hinsichtlich der Klimadimensionen entwickelt (Eder, 2001; Pekrun, 1985). Der Fragebogen mit Aussagen zum Unterricht, die von den Schülern anhand eines bestimmten Antwortformats danach zu beurteilen sind, wie sehr sie auf die Klasse bzw. den Unterricht zutreffen, ist bei den Klimaforschern die Methode der Wahl (Dreesmann, 1982).

Saldern und Littig (1987) entwickelten die *Landauer Skalen zum Sozialklima* (LASSO), die neben der (1) Lehrer-Schüler- und (2) Schüler-Schüler-Beziehungen auch (3) allgemeine Merkmale des Unterrichts in Form von Leistungsdruck, Zufriedenheit mit dem Unterricht, Disziplin und Ordnung, Fähigkeit des Lehrers zur Vermittlung von Lerninhalten, Resignation sowie eine reduzierte Unterrichtsteilnahme erfassen (Beiheft, S. 3). Eder (Eder, 1996) erweiterte diese Dreiteilung um die Dimension der Lernhaltungen des Schülers. Die weiteste Verbreitung im Forschungsbereich haben der *Klassenklima-Fragebogen* von Walberg (1966) und der *Learning Environment Inventory* (Anderson, 1973, zitiert nach; Kemme, 1998). Die Arbeiten von Moos (Moos & Trickett, 1974) erfassen u. a. auch Regelklarheit, Kontrolle und Aufgabenorientierung, die Eder (2001) den Unterrichtsmerkmalen im Sinne der Lehr-Lernforschung zugehörig sieht. Einschlägige Übersichten zu den Dimensionen dieser und weiterer Verfahren finden sich bei Gruehn (2000), Clausen et al. (2003), Schröder (1999) und Eder (1996).

Die Fragebogenmethode bringt ökonomische Vorteile, aber auch methodisch-instrumentelle Messprobleme mit sich, die u. a. aus der Konstruktheterogenität der Unterrichtsqualität (s. Abschnitt 2.3) sowie des spezifischen Kontexts der Schülerbefragung resultieren. Im Folgenden sollen die Itemformulierung, das Antwortformat und der Urteilsbezug als beeinflussende Faktoren der Reliabilität und Validität von Schülerwahrnehmungen betrachtet werden.

### 3.4.1 Itemformulierung

Eine wichtige Betrachtungsdimension hinsichtlich der Itemformulierung ist der *evaluative Gehalt* des Items (Eder, 2002; 2001), der sich in der Wertigkeit des Items manifestiert und anhand seiner Extrempole ‚wertend‘ (z. B. „Mit unserem Lehrer sind wir sehr zufrieden“, Item 20, LASSO 4-13, Saldern & Littig, 1987) bis hin zu ‚beschreibend‘ (z. B. „In der Stunde war es so ruhig, dass ich gut arbeiten konnte“, Item 7, SFB) verdeutlichen lässt. Diese Dimension hängt auch mit der Beobachtbarkeit und der Verhaltensnähe zusammen, die wiederum mit der Abstraktionsebene bzw. dem *Inferenzgrad* kontaminiert sind. Niedrig-inferente Items (z. B. „Die Stunde hat pünktlich begonnen“, Item 1, SFB) thematisieren beobachtbare Verhaltensaspekte, die aufgrund ihrer Spezifität relativ ‚objektiv‘ erfasst werden können (Clausen et al., 2003). Kognitive Schlussfolgerungen oder Interpretationen über die konkret beobachtbare Verhaltenssequenz hinaus sind auf Beobachterseite kaum nötig, daher lässt sich bei Verwendung dieses Itemtyps eine relativ hohe Übereinstimmung

innerhalb der Klasse feststellen (Dreesmann, 1982). Hoch-inferente Items (z. B. „Unser Lehrer glaubt, Pünktlichkeit und Ordentlichkeit sind das Wichtigste“, Item 116, LASSO 4-13, Saldern & Littig, 1987) erfordern hingegen „Schlussfolgerungen bzw. interpretative Prozesse seitens der Beobachter, die über das konkret beobachtbare Verhalten hinausgehen und sich auf abstraktere Sachverhalte bzw. globalere Verhaltensmerkmale beziehen“ (Clausen et al., 2003, S. 124). In diesem Fall nimmt der Schüler nicht mehr eine Beobachter- sondern eine Beurteilerrolle ein (vgl. Dreesmann, 1982), die mit den in Abschnitt 3.3 beschriebenen potentiellen (un-)systematischen Beurteilungsverzerrungen einhergeht. Dreesmann (1982) rät zur Mischung von Items unterschiedlicher Abstraktionsgrade, um beobachtbare und nicht-beobachtbare Unterrichtsaspekte abzudecken. Damit umgeht er gleichzeitig das „Übereinstimmungs-Bedeutsamkeits-Dilemma“ (Clausen, 2002, S. 191), welches besagt, dass für die Einschätzung niedrig-inferenter Aussagen als ‚objektivere Variante‘ zwar wenig pädagogisch-didaktisches Verständnis notwendig und das Ausmaß von Beurteilungsfehlern eher gering ist, hoch-inferente Aussagen aber höhere Validitätskoeffizienten in Bezug auf schulische Leistungs- und Entwicklungskriterien aufweisen (Fraser & Walberg, 1981, zitiert nach Clausen et al., 2003).

Ein weiterer Faktor ist die *Wahrnehmungsperspektive* – Eder (2002) unterscheidet zwischen „ich über mich“, „wir über uns [Klasse]“ oder „wir über sie [die Lehrer]“ (S. 219). Wagner (2008) differenziert unter Rückgriff auf den Brok, Brekelmans und Wubbels (2006), wie in Tabelle 3 dargestellt, hinsichtlich der Klasse vs. dem individuellen Schüler als Adressaten des Lehrerverhaltens und operationalisiert die Wahrnehmungsperspektive als Gruppenwahrnehmung vs. personalisierte Wahrnehmung.

Tabelle 3: 4-Felder-Matrix zur Wahrnehmungsperspektive vs. Adressat des Lehrerverhaltens sowie der Zuordnung von Eders Klassifikation (2002, S. 219) anhand des Items 21 im Projektfragebogen (in Anlehnung an Wagner, 2008, S. 9)

		Adressat des Lehrerverhaltens	
		individueller Schüler	Klasse
Wahrnehmungsperspektive	individuelle Sicht	<b>„Ich finde, mit Fehlern ist die Lehrkraft so umgegangen, dass ich daraus etwas lernen konnte“</b> Eder (2002): „ich über mich“	„Ich finde, die Lehrkraft ist mit Fehlern so umgegangen, dass wir daraus etwas lernen konnten“
	Klassensicht	„Wir finden, die Lehrkraft ist mit Fehlern so umgegangen, dass ich daraus etwas lernen konnte“	<b>„Wir finden, mit Fehlern ist die Lehrkraft so umgegangen, dass wir daraus etwas lernen konnten“</b> Eder (2002): „wir über uns“

Anmerkung: Die Hervorhebungen beziehen sich auf die im Projektfragebogen verwendeten Items.

Die Itembeispiele legen nahe, dass theoretisch immer dann, wenn eine klassenbezogene Beurteilung erfolgt, sei es hinsichtlich der Resultate auf die Klasse gerichteten Lehrerverhaltens und/oder die Einschätzung dieser Sachverhalte aus Klassensicht, sowohl die Kenntnis bezüglich des Urteilsgegenstands als auch eine Aggregation der entsprechenden Informationen nötig sind. Soll z. B. der an die Klasse adressierte Umgang des Lehrers mit Schülerfehlern beurteilt werden, muss bekannt sein, ob die Mitschüler einen Lerneffekt aufgrund der Lehrerreaktion hatten. Im Einzelfall mögen Lernerfahrungen unter Schülern ausgetauscht werden, solche reflexiven Gespräche stellen aber eher die Ausnahme dar. Die Erfahrungen der einzelnen Mitschüler müssten anschließend in einen ‚Klasseneffekt‘ integriert werden. Wird dieser Sachverhalt aus der Klassenperspektive beurteilt, bedürfte es neben der Information, inwiefern die Mitschüler vom Lehrerverhalten profitiert haben, auch der Einschätzung der Mitschüler zur Lernerfahrung. Auch in diesem Fall würde die Aggregation der Einschätzungen zu einer ‚wir finden‘-Klassenmeinung folgen. Diese Aggregation ist in vielfacher Hinsicht als problematisch zu erachten: (1) Die Meinungen bzw. Einschätzungen aller Mitschüler sind in der Regel nicht bekannt, (2) das Aggregieren ist ein kognitiv aufwendiger Prozess und (3) das Aggregat könnte im Widerspruch zum eigenen Erleben bzw. der eigenen Meinung stehen, d. h. konnte ein Schüler als Einziger die Lehrerreaktion nach einem Fehler nicht in einen Lerneffekt überführen, weiß aber, dass dies den anderen gelungen ist, müsste er der Aussage „Mit Fehlern ist die Lehrkraft so umgegangen, dass wir daraus etwas lernen konnten“ (Item 21, SFB) aus beiden Wahrnehmungsperspektiven zustimmen, obwohl es der eigenen Einschätzung widerspricht. Hinsichtlich der ersten beiden Aspekte ist aus der ‚survey and cognition‘-Forschung (Sudman, Bradburn & Schwarz, 1996) übertragbar, dass der Suchprozess nach relevanten Informationen bei der Urteilsbildung ‚on the spot‘ in der Regel abgebrochen wird, wenn genügend Informationen für ein Urteil abgerufen werden konnten (Taylor, 1981). Dabei kommen Heuristiken zum Tragen, indem z. B. auf die Meinung der klasseninternen Peer-Gruppe bzw. Clique, die des Meinungsführers in der Klasse oder eines ‚repräsentativen‘ Mitschülers zurückgegriffen und anschließend eine Gewichtung im Vergleich zur eigenen Position vorgenommen wird. Je nach Bezugsmaßstab kann es dabei zu Assimilations- oder Kontrasteffekten kommen, d. h. zur Addition von Informationen gleicher oder gegenteiliger Valenz (Sudman et al., 1996). Vor dem Hintergrund dieser sozialpsychologischen Befunde und angesichts der hohen Anzahl an Fragebogenitems sowie weiterer Dimensionen, die im Urteilsprozess beachtet werden müssen, z. B. der Auswahl eines repräsentativen Verhaltensausschnitts des

Lehrers und einem potentiellen interindividuellen Vergleich zwischen verschiedenen Lehrkräften, erscheinen Urteile aus der wir-Perspektive daher als äußerst komplex.

Die Beurteilung auf den beurteilenden Schüler selbst gerichteten Lehrerverhaltens aus seiner eigenen Perspektive weist die eben genannten Probleme nicht auf und scheint nur geringe kognitive Anforderungen zu stellen, allerdings bleiben auch hier Messfehler aufgrund psychologischer Beurteilungsfehler, die allerdings auch bei allen anderen Itemtypen auftreten können. Wird allerdings klassenbezogenes Lehrerverhalten aus individueller Sicht beurteilt (oder individuenzentriertes Lehrerverhalten aus Klassensicht), fordert die mangelnde Symmetrie im Urteilsbezug zusätzliche Verarbeitungskapazitäten. Liegt nämlich eine auf Klassenebene aggregierte Einschätzung des Iteminhalts bzw. eine repräsentative Klassenmeinung vor, muss diese mit der eigenen Meinung bzw. Erfahrung noch zusätzlich verglichen werden, um ein abschließendes Urteil bezüglich der Antwortformierung treffen zu können.

### **3.4.2 Antwortformat**

In der Regel kommen geschlossene Fragen in schriftlicher Form zur Anwendung, mit Hilfe derer die Schülerzustimmung hinsichtlich des Zutreffens der Itemsaussage erfasst wird (vgl. Dreesmann, 1982). Das offene Frageformat wird aufgrund des erheblichen Auswertungsaufwands und der fehlenden Vergleichbarkeit der Ergebnisse relativ selten eingesetzt. Hinsichtlich der Antwortskala zeigen Follman, Lucoff, Small und Power (1974), dass sich die Itemmittelwerte bei Professorenbeurteilungen durch Studenten in Abhängigkeit von den Skalenbeschreibungen zwar in der absoluten Höhe unterscheiden, aber nicht signifikant variieren. Absolut gesehen sind die Itemmittelwerte bei einer evaluativen Skala (unbefriedigend bis exzellent) geringer als bei einer Zustimmungsskala (stimme sehr zu bis stimme absolut nicht zu) bzw. bei einer Skala, die den Verbesserungsbedarf erfasst (hoher Verbesserungsbedarf bis kein Verbesserungsbedarf). Preston und Colman (2000) können bezüglich der Anzahl an Antwortalternativen zeigen, dass erst Antwortskalen ab fünf Antwortkategorien akzeptable Skalenreliabilitäten aufweisen. Vor dem Hintergrund des Fragebo-geneinsatzes im Unterricht ist es allerdings auch wichtig, die Bearbeitungsschwierigkeit und die damit einhergehende Bearbeitungszeit zu berücksichtigen (vgl. Wagner, 2008).

### 3.4.3 Urteilsbezug

Das Fragebogenverfahren weist einen retrospektiven Selbstberichtscharakter auf. Dies kann u. a. mit Erinnerungsverzerrungen einhergehen. Der Bezug auf eine konkrete, z. B. kürzlich stattgefundene Unterrichtsstunde, weist den Vorteil der Reduktion potentieller Verzerrungseffekte auf, die aus der Rekonstruktion von Verhaltenssequenzen resultieren. Nach Sudman et al. (1996) beruhen Einstellungsurteile auf Informationen, die leicht verfügbar sind, während andere potentiell relevante Informationen geringer Verfügbarkeit mit hoher Wahrscheinlichkeit nicht beachtet werden. Die Verfügbarkeit ist dabei eine Funktion der (1) ‚recency‘, welche die zeitliche Dimension in dem Sinne erfasst, dass vor kurzem genutzte Informationen eher zugänglich sind als solche, die vor längerer Zeit verwendet wurden und (2) der ‚frequency‘ als Nutzungshäufigkeit der Information, d. h. eine chronisch verfügbare Information wäre im Unterrichtskontext, dass ein bestimmter Schüler regelmäßig zu spät zum Unterricht kommt und bei allen Mitschülern dafür bekannt ist (vgl. Bless et al., 2004). Beziehen sich Unterrichtswahrnehmungen auf eine kürzlich gehaltene Unterrichtsstunde, können ‚recency-Effekte‘ der Informationsverarbeitung und den Urteilsprozessen angenommen werden, d. h. die zur Beurteilung herangezogenen Informationen weisen mit großer Wahrscheinlichkeit eine hohe Schnittmenge mit den in der konkreten Unterrichtsstunde erlebten Verhaltenssequenzen auf.

Die personenbezogene vs. fachbezogene Urteilsdimension charakterisiert, inwiefern die Beurteilung auf eine bestimmte Lehrkraft vs. ein bestimmtes Fach abzielt oder die Einschätzung der Lehrkräfte im Allgemeinen vs. fächerübergreifende Einschätzungen anstrebt. Letztere erfordern eine Verallgemeinerung über die Zeit vs. verschiedene Fächer (vgl. Eder, 2002), die mit den im Abschnitt 3.4.1 beschriebenen Aggregierungsproblemen einhergeht. Ditton (2002b) weist auch darauf hin, dass eine fächerübergreifende Einschätzung aufgrund der Mittlung über verschiedene Lehrkräfte zu einer erschwerten Interpretierbarkeit führt.

Zusammenfassend lässt sich sagen, dass die Itemformulierung und -konstruktion gerade im Unterrichtsentwicklungskontext eine hohe Bedeutung hat. Schüler, wie Menschen im Allgemeinen, verfügen über eine begrenzte Verarbeitungskapazität, wird diese in hohem Maße beansprucht, z. B. aufgrund komplexer Itemformulierungen, bei der mehrere Urteilspektiven eingenommen und miteinander verglichen werden müssen, erhöht sich die Wahrscheinlichkeit einer konzeptgesteuerten Informationsverarbeitung (Bless & Schwarz, 2002). Diese geht u. a. mit einem verstärkten Rückgriff auf passendes und verfügbares

Vorwissen einher, neu eintreffenden Stimulusinformationen werden weniger Beachtung geschenkt. Darunter würden auch subtile Änderungen im Lehrerverhalten im Schülerfeedbackprozess fallen, die nicht in das Urteil bei einer Wiederholungsmessung eingingen. Die Verarbeitungsintensität ließe sich durch klare und hinsichtlich der Wahrnehmungsperspektive eindimensionale Itemformulierungen vereinfachen sowie einer hohen Verarbeitungsmotivation als Folge hoher Relevanz der Thematik. Das Erfassen der Schülerangaben allein reicht jedoch nicht aus, wichtig ist auch, dass diese zur Unterrichtsentwicklung genutzt werden. Wie dies im pädagogischen Alltag gelingen kann, wird im nächsten Kapitel erörtert.

## **4 Schülerfeedback als Beitrag zur Unterrichtsentwicklung**

Unterrichtswahrnehmungen von Schülern als Selbstevaluationsinstrument bieten ein hohes Potential zur Unterrichtsentwicklung. In diesem Kapitel wird das Schülerfeedbackverfahren daher bezüglich seiner Rolle als internes Instrument der Selbstevaluation (Abschnitt 4.1), hinsichtlich relevanter Wirksamkeitsdeterminanten (Abschnitt 4.2) und in Bezug auf seine Implementierung in den pädagogischen Alltag beleuchtet (Abschnitt 4.3).

### **4.1 Evaluation als Funktion von Schülerfeedback**

Rindermann versteht unter Evaluation die „systematische Analyse und empirische Untersuchung von Konzepten, Bedingungen, Prozessen und Wirkungen zielgerichteter Aktivitäten zum Zwecke ihrer Bewertung und Modifikation“ (2001, S. 61; vgl. Wottawa & Thierau, 2003). Eine Evaluation ist durch eine nutzenorientierte Perspektive gekennzeichnet, d. h. die typische Frage des Evaluators ist, ob der zu evaluierende Gegenstand seinen Zweck erfüllt und/oder welche Veränderungen erforderlich sind, um seinen Zweck noch besser zu erfüllen (Kromrey, 2001). Schülerfeedback erfüllt die konstituierenden Merkmale einer Evaluation, wenn dabei systematisch Informationen zum Unterricht eingeholt und einer nutzenorientierten Bewertung unterzogen werden. Der Nutzen besteht dabei in erster Linie in der Reflexion des pädagogischen Handelns in der Lehrerrolle zur Weiterentwicklung der eigenen Professionalität und somit der Verbesserung der Lehr-Lernprozesse. Gegenüber einer gemäß Altrichter (1999) ohnehin im pädagogischen Alltag stattfindenden Reflexion eigenen Handelns, grenzt sich das Schülerfeedbackverfahren durch die systematische und zielorientierte Vorgehensweise ab. Die dafür notwendigen Implementierungsschritte sind Gegenstand des Abschnitts 4.3.



Da beim Schülerfeedback das Evaluationsteam und die betroffenen Stakeholder derselben Organisationseinheit angehören, kann dieses als *interne Evaluationsmaßnahme* kategorisiert werden – im Gegensatz zur externen Evaluation, bei der das Evaluationsteam organisationsfremd ist (vgl. Hense, 2006; Widmer, 2000). Die klassische Rollentrennung zwischen den Auftraggebern, den Evaluierten und Evaluierenden wird im Kontext der Schülerbefragung durch eine Partizipation der Betroffenen abgelöst (vgl. Hense, 2006). Die Lehrkraft nimmt sowohl die Auftraggeber- als auch zusammen mit den Schülern als Datenlieferanten die Evaluatoren- und die Stakeholderrolle ein. Diese Rollenidentität, d. h. dass „die praxisgestaltenden Akteure identisch sind mit den evaluierenden Akteuren“ (Müller-Kohlenberg, 2004, S. 5), ist ein konstituierendes Merkmal der *Selbstevaluation*. Die Lehrkräfte beauftragen in ihrer Auftraggeberrolle die Schülerbefragung, ‚finanzieren‘ diese v. a. durch zeitlichen Aufwand in Form von Vorbereitungs- und Unterrichtszeit und sind die Hauptadressaten des Ergebnisberichts. Als Evaluatoren übernehmen sie die Verantwortung für die Planung und Durchführung der Schülerbefragung und sind gleichzeitig als betroffene Stakeholder für die Konzeption und Umsetzung des Unterrichts als Evaluationsgegenstand verantwortlich. Die Schüler sind als Feedbackgeber und Reflexionspartner ebenfalls Teil des Evaluationsteams und gleichzeitig Stakeholder, die von den aus der Evaluation resultierenden Konsequenzen profitieren sollen.

Selbstevaluation besagt allerdings nicht, dass sämtliche zum Evaluationsprozess gehörenden Tätigkeiten von den Praktikern, d. h. im vorliegenden Kontext den Lehrkräften, übernommen werden müssen. Es genügt, wenn im Evaluationsprozess wesentliche Schritte selbstverantwortlich durchgeführt werden (vgl. Hense, 2006). Die Methodik der Selbstevaluation schließt dabei die Unterstützung und Ergänzung durch externe Evaluationsmaßnahmen nicht aus, sondern wird im Sinne des Aufdeckens ‚blinder Flecken‘ als Korrektiv sogar empfohlen (Burkhard, 1997; Scriven, 1991). Pikowsky (2007a) rät deshalb hinsichtlich der Qualifizierung externer Schulevaluationsteams, „...die Anschlussfähigkeit der externen Evaluation an die interne Evaluation der Qualitätsprogramme sowie die Qualitätsarbeit der Schulen“ (S. 1) als wichtiges Ziel zu fokussieren.

Der partizipative Charakter des Schülerfeedbacks bringt *Chancen und Risiken* mit sich. Dadurch, dass die wesentlichen Planungs- und Durchführungsschritte im Evaluationsprozess in einer und gleichsam der Hand des Praktikers liegen, sollte die Akzeptanz größer sein und weniger Befürchtungen, Konflikte, Widerstände sowie Evaluationsängste ausge-

löst werden (Buhren, Killus & Müller, 2000; Burkhard & Eikenbusch, 2000; Dorenkamp, 1999). Ferner sind die Lehrkräfte und Schüler als ‚Betroffene‘ mit dem Evaluationsgegenstand sehr gut vertraut, kennen informelle Abläufe und das Evaluationsumfeld (s. dazu auch Abschnitt 3.3). Die Wahrscheinlichkeit der Ableitung und Umsetzung von Konsequenzen ist aufgrund der Verantwortung für den Evaluationsprozess, der intensiven Auseinandersetzung mit dem Evaluationsinstrument und -gegenstand während der Evaluationsphase und der möglichen unmittelbaren Rückspeisung der Evaluationsergebnisse in die Praxis erhöht (Greene, 1988). Selbstevaluation hilft somit, den Evaluationsnutzen zu steigern (O'Sullivan & D'Agostino, 2002). Auf der anderen Seite bringt die Selbstevaluation die Problematik einer potentiell mangelnden Objektivität, Parteilichkeit und Beschönigungsgefahr hinsichtlich der Ergebnisbeschreibung und -interpretation mit sich (Widmer, 2000). In der Regel weisen Lehrkräfte außerdem kaum Evaluationskenntnisse und nur eine geringe Qualifizierung hinsichtlich der Evaluationssteuerung auf.

Werden die in der Literatur zahlreichen Auflistungen zu den möglichen Zielen einer Evaluation (Wesseler, 1999; Westermann, 2002; Will, Winteler & Krapp, 1987; Wulf, 1975a; Wulf, 1975b) vor dem Hintergrund des Schülerfeedbacks inspiziert, lassen sich, bei einem Verständnis der Evaluation als inhärent bewertungsorientiert, zwei *Primärfunktionen* identifizieren: Die *Entwicklungsfunktion* im Sinne der Verbesserung bzw. Optimierung der Lehr-Lernprozesse als evaluierten Gegenstand und die *Lernfunktion*, welche sich v. a. in einem Erkenntnisgewinn in Bezug auf den Evaluationsgegenstand und den -prozess an sich äußert. Hense (2006) ordnet die Entwicklungsfunktion, wie in Tabelle 4 dargestellt, anhand der Dimensionen (1) Prozess als systematische Untersuchung des Evaluationsgegenstands vs. Produkt in Form der Bereitstellung konkreter Evaluationsergebnisse und (2) Handlungsebene als konkrete Veränderungen hinsichtlich des Evaluationsgegenstands vs. Wissenssebene in Bezug auf die bei der Evaluation erworbenen Informationen und Erkenntnisse über den evaluierten Gegenstand jeweils den beiden ersten Dimensionspolen zu. Schülerfeedback trägt somit in seiner Entwicklungsfunktion dazu bei, konkrete Schwachstellen bzw. Ist-Soll-Diskrepanzen aus der Sicht des Schülers aufzudecken (Prozessnutzen) und konkrete Veränderungen zu begründen bzw. auszulösen (Handlungsebene). Dies wird auch über die Lernfunktion mediert, die ermöglicht, mehr über die Wirkungsweise des evaluierten Gegenstands zu erfahren (Produktnutzen) und Veränderungen hinsichtlich der Einstellungen, Überzeugungen sowie einen Wissenszuwachs bei den beteiligten Evaluatoren zu bewirken (Wissensebene).

Tabelle 4: *Hauptfunktionen der Evaluation (in Anlehnung an Hense, 2006, S. 51)*

	Prozessnutzen	Produktnutzen
Handlungsebene	(1) Entwicklungsfunktion	(2) Entscheidungsfunktion
Wissensebene	(3) Legitimationsfunktion	(4) Lernfunktion

Eine Entscheidungs- und Legitimationsfunktion können Selbstevaluationsergebnisse im Evaluationskreislauf der Bestimmung des Ist-Zustands, der in Abhängigkeit von der Ist-Soll-Diskrepanz durchgeführten Interventionsmaßnahmen und dem erneuten Erfassen des aktuellen Zustandes einnehmen. Dabei wird gemäß der Selbststeuerung der Lehrkraft lediglich auf selbst gewählte Legitimationen und Entscheidungen im Rahmen einer professionellen Selbstkontrolle Bezug genommen (Buhren et al., 2000; König, 2000). Zur Rechtfertigung von Entscheidungen durch Akteure, die sich außerhalb des Selbstevaluationsprozesses befinden, sind Selbstevaluationsergebnisse aufgrund von außen oft kritisch betrachteter Vollständigkeits-, Objektivitäts- und Glaubwürdigkeitsaspekte ohnehin eher ungeeignet. Da eine Selbstevaluation eher nach innen gerichtet und prozessorientiert ist – im Vergleich zur Fremdevaluation, bei der nach Widmer (2000) der Fokus stärker auf der Rechenschaftslegung mit Produktcharakter liegt – eignet sich diese v. a. bei formativ ausgerichteten Evaluationen (Scriven, 1991). Bei diesen erfolgt die Evaluation parallel zur Maßnahmendurchführung bzw. -implementierung und kann somit in erster Linie eine kontinuierliche (Unterrichts-)Entwicklung und auch Evaluationsprozessoptimierung über die Zeit fokussieren. Eine ausgeprägte Kenntnis des zu evaluierenden Gegenstands ist Voraussetzung formativer Evaluationen und genau dieses Wissen bringen Selbstevaluatoren mit (Wottawa & Thierau, 2003).

Das Schülerfeedback weist ferner eine Reihe von *Sekundärfunktionen* auf verschiedenen Ebenen auf und kann seine Wirkung auch ebenen-übergreifend in dem Sinne zeigen, dass die Verbesserung der Einzelleistung einer Systemteilkomponente in seiner Summe zur Optimierung des Gesamtsystems führt. Daumen, Miethner und Pikowsky (2005; zitiert nach Pikowsky, 2007b) betonen diesbezüglich in Anlehnung an Kempfert und Rolff (2005), dass die schulische Qualitätsentwicklung eine Funktion der Unterrichts-, Personal- und Organisationsentwicklung sein sollte, auf deren Ebenen wiederum die allgemeinen Voraussetzungen, die ablaufenden Arbeitsprozesse und Wirkungen zu fokussieren sind. Ein wichtiger Sekundäreffekt, der v. a. auf der Personalentwicklungsebene anzusiedeln ist, stellt die Weiterbildungsfunktion dar, die sich in der Metapher ‚evaluator as educator‘ von

Cronbach et al. (1980) ausdrückt. Sie umschreibt die nahezu beiläufig stattfindende Weiterqualifizierung des Evaluators durch die Reflexion der Evaluationsprozesse und -ergebnisse (vgl. Hense, 2006). Im Kontext des Schülerfeedbackverfahrens erhält die Lehrkraft z. B. ein Verständnis der Operationalisierung lernwirksamer Unterrichtsmerkmale. Außerdem werden Reflexionen bezüglich der Unterrichtsziele und schulischer Qualitätskriterien angeregt. Diese Prozesse stärken die Eigenverantwortung der Lehrkräfte und ermöglichen deren stärkere Einbindung in Steuerungsprozesse bei Schnittstellenthematiken im Rahmen einer gestärkten Autonomie der Einzelschule. Ferner weist die Selbstevaluation ein großes Potential im Sinne der Verbesserung der Kommunikations- und Kooperationsstrukturen durch den verstärkten Dialog zwischen der Lehrkraft als Evaluator und den Schülern als Evaluierenden bzw. (Fach-)Kollegen und/oder der Schulleitung sowie den Eltern als potentielle Reflexionspartner auf (MacBeath, Schratz, Meuret & Jakobsen, 2000). Die Selbstevaluation bringt aber auch für die Schüler als Feedbackgeber eine Reihe denkbarer Vorteile mit sich. Dazu gehören das Potential zur Schulung der Selbstwahrnehmung und -reflexion, v. a. durch Fragebogenitems mit einem ich-Bezug und zur Verbesserung der Kommunikations- und Konfliktfähigkeiten durch den verstärkten Schüler-Schüler- und Schüler-Lehrer-Dialog. Und schließlich ergibt sich die Möglichkeit, am Modell der Lehrkraft zu lernen, wie Entwicklungsherausforderungen angenommen und gelöst werden können (vgl. Bandura, 1976). Gleichzeitig fungiert Schülerfeedback auch als Teilelement der Unterrichtsmerkmale Schüler- und Wirkungsorientierung (s. Abschnitt 2.2.4) als abhängige Variable, indem durch das Einholen des Feedbacks die Schülermeinungen erfasst und ernst genommen werden. Die Selbstevaluation stellt somit ein sehr gutes Verbindungsglied zwischen der Unterrichts- und Personalentwicklung mit einem fließenden Übergang zur Organisationsentwicklung dar. Welche Faktoren die positiven Wirkungen des Schülerfeedbacks im Kontext der Selbstevaluation determinieren, wird im folgenden Abschnitt betrachtet.

## **4.2 Determinanten der Wirksamkeit von Selbstevaluation**

Inwiefern Selbstevaluationsergebnisse zur Unterrichtsqualität eine wirksame Stellschraube zur Optimierung und Verbesserung unterrichtlicher Prozesse sind, hängt von zahlreichen Determinanten ab. Zwei grundlegende Voraussetzungen sind: (1) Es muss ein kausaler Zusammenhang zwischen der Unterrichtsqualität und den intendierten Unterrichtszielen bestehen. Für kognitive und motivationale Kriterien wurde dieser Nachweis im Abschnitt 2.1 und 2.2.4 geführt. (2) Unterrichtsqualität muss veränderbar sein. Der Unterrichtsqualität

kommt diesbezüglich im komplexen Gefüge verschiedener Schulleistungsdeterminanten eine besondere Bedeutung zu. Im Vergleich zur Intelligenz der Schüler oder den Persönlichkeitsmerkmalen der Lehrkraft wird die Unterrichtsqualität aufgrund der über das Lehrer- und Schülerverhalten mediierten situationsabhängigen State-Anteile als stärker veränderbar angesehen (Einsiedler, 1997). Neben diesen allgemeinen Voraussetzungen gibt es eine Reihe von theoretischen Ausführungen zu konkreten distalen und proximalen Determinanten der Wirksamkeit von Selbstevaluationen im Allgemeinen, deren empirische Überprüfung bis auf wenige Ausnahmen noch aussteht. Unter Bezug auf das Zyklenmodell von Helmke und das Rahmenmodell zu den Wirkungen und Erfolgsfaktoren der Selbstevaluation werden im Folgenden wesentliche Determinanten vorgestellt. *Helmke* (2006a, S. 200; Helmke, 2007c, S. 14; Helmke & Hosenfeld, 2005, S. 147) postuliert in seinem Zyklenmodell zur Unterrichtsentwicklung, dass ausgehend von den in einer internen (oder externen) Evaluation erhobenen (1) unterrichtsbezogenen Daten die Prozesse der (2) Datenrezeption und (3) -reflexion zu durchlaufen sind, um schließlich konkrete Maßnahmen in der (4) Aktionsphase umsetzen und anschließend evaluieren zu können. Dieser Prozess wird durch eine Reihe von individuellen und externen Bedingungen beeinflusst. Zu den individuellen Bedingungen der an der Unterrichtsentwicklung unmittelbar Beteiligten gehören u. a. das Selbstkonzept, der subjektive Leidensdruck, die Bereitschaft zur Selbstreflexion und Kooperation, ein professionelles Selbstverständnis, die persönliche Bedeutsamkeit des eigenen Unterrichtserfolgs, die Plastizität von Gewohnheiten sowie Routinen, die Akzeptanz von Evaluationen und die Selbstwirksamkeit. *Balk* (2000) zeigt für den Kontext der Lehrevaluation, dass die Feedbackverwertung bei hoch ängstlichen Personen und Personen mit einer geringen Selbstwirksamkeit weniger gut gelingt. Schulinterne Faktoren sind gemäß *Helmke* z. B. ein kooperatives Klima im Kollegium, die vorherrschende Evaluations- und Innovationskultur, die Unterstützung durch die Schulleitung sowie das schulinterne Fortbildungs- und Beratungsangebot. Diese bilden zusammen mit den schulexternen Kontextfaktoren, wie z. B. der Wertschätzung und Unterstützung durch die Schulaufsicht, die Akzeptanz durch die Eltern und die Schülerschaft sowie Hilfeleistungen durch die Wissenschaft die externen Bedingungen des Unterrichtsentwicklungsprozesses.

*Hense* (2006, S. 149) führt eine Reihe von Input- (u. a. Erwartungen, Befürchtungen, Vertrautheit mit dem Verfahren, dispositionelle Voraussetzungen, Vorabinformation der Beteiligten, organisationale Anbindung) und Prozesselementen (u. a. Selbstbestimmung, Engagement, wissenschaftliche Begleitung) als Erfolgsfaktoren der Wirkung bzw. des Outcomes

der Selbstevaluation an. Dabei wird über alle drei Stufen hinweg in Übereinstimmung mit dem Postulat von Helmke zwischen der individuellen und überindividuellen Ebene, welche die Rahmenbedingungen und Organisationsmerkmale der Selbstevaluation erfassen, unterschieden. Im Rahmen eines Programms zur systematischen Einbeziehung von Medien, Informations- und Kommunikationstechnologien in Lehr- und Lernprozessen (SEMIK) befragt Hense (2006) wissenschaftlich begleitete Selbstevaluationsgruppen (Lehrkräfte verschiedener Schulen) hinsichtlich dieser Erfolgsfaktoren als Prädiktoren für die anhand von drei abhängigen Variablen operationalisierten Selbstevaluationswirkungen: (1) Der persönliche Nutzen, welcher anhand von Items zur verstärkten Selbstreflexion und Weiterqualifizierung erfasst wird, (2) der überindividuelle Nutzen, welcher über Items zum Erkennen von Stärken und Schwächen, Kommunikationsverbesserungen sowie Professionalisierungen der eigenen Tätigkeit operationalisiert wird und (3) das Kosten-Nutzen-Verhältnis zur Erfassung des investierten Aufwands im Verhältnis zum Selbstevaluationsnutzen erfasst. Regressionsanalytische Untersuchungen zeigen, dass die retrospektiv erfassten *Erwartungen* bezüglich positiver Wirkungen der Selbstevaluation signifikant zur Vorhersage des individuellen und überindividuellen Nutzens beitragen, das Kosten-Nutzen-Verhältnis wird negativ durch *Befürchtungen* arbeitsintensiver oder ineffektiver Selbstevaluationsmaßnahmen beeinflusst. Im Gegensatz zum persönlichen Nutzen, der neben den bereits erwähnten Prädiktoren durch ein *kooperatives Vorgehen* und eine *wissenschaftliche Begleitung* determiniert wird, hängt der projektbezogene Erfolg eher von der *Vorabinformation* der Beteiligten ab. Die *Selbstbestimmung*, die sich auf die eigenverantwortliche Selbststeuerung des Evaluationsprozesses und die Verfügungsgewalt über die Ergebnisse bezieht, stellt für alle abhängigen Variablen einen signifikanten Prädiktor dar und nimmt damit einen entscheidenden Erfolgsfaktor ein. Die inhaltliche und quantitative Passung der *Instrumente* der Selbstevaluation und eine praxisnahe und fachlich hochwertige *wissenschaftliche Begleitung* gehen mit einer hohen Kosten-Nutzen-Relation einher, wobei letztere auch einen signifikant positiven Einfluss auf den individuellen Nutzen ausübt. Hense (2006) weist darauf hin, dass dies ein Hinweis für die Relevanz der notwendigen Evaluationsexpertise bei der Selbstevaluation ist.

Eine wichtige Rolle spielt auch die *Gestaltung der Ergebnissrückmeldung*, denn gemäß Helmkes Zyklenmodell fungieren die den Lehrkräften rückgemeldeten Informationen gleichsam als Input für die nachfolgenden Prozesse der Rezeption und Reflexion. Die Frage danach, was eine Rückmeldung von Schülerfeedback konkret beinhaltet und wie diese

gestaltet sein sollte, um bei der Lehrkraft Veränderungsbereitschaft zu wecken sowie Veränderungsprozesse und einen fruchtbaren Dialog mit den Schülerinnen und Schülern zu initiieren, bedürfte genau genommen einer eigenen umfassenden wissenschaftlichen Untersuchung. Deshalb soll an dieser Stelle auf die Dissertation von Schneewind (2007) zu dieser Fragestellung im Kontext von Schulleistungsstudien verwiesen werden. Ausgehend von den bisher vorgestellten Determinanten der Selbstevaluation im Allgemeinen lassen sich jedoch einige Überlegungen darüber festhalten, wie rezeptions- und reflexionsdienliche Ergebnisrückmeldungen beschaffen sein sollten: (1) Ergebnisrückmeldungen sollten zeitnah erfolgen, einfach handhabbar und verständlich sein, d. h. nur relevante Informationen und notwendige elementare statistische Grundbegriffe, die zielgruppengerecht erläutert werden, enthalten. (2) Schülerfeedback sollte auf Klassenniveau aggregiert rückgemeldet werden, um die Anonymität des einzelnen Schülers zu gewährleisten. (3) Die Ergebnisse sollten auch auf Itemebene rückgemeldet werden, da aggregierte Skalenwerte einen Informationsverlust bedeuten und hinsichtlich der angestrebten Veränderungsprozesse keine ausreichend konkreten Ansatzpunkte auf Verhaltensebene liefern.

Die Befunde zu den Wirksamkeitsdeterminanten und die im Abschnitt 4.1 erörterten allgemeinen Aspekte der Selbstevaluation werden im nächsten Abschnitt hinsichtlich der Konsequenzen für die Implementierung des Schülerfeedbackverfahrens auf Unterrichts-, aber auch auf Personal- bzw. Organisationsentwicklungsebene zusammengeführt.

### **4.3 Implementierung des Schülerfeedbackverfahrens**

Wie das Schülerfeedbackverfahren von der Lehrkraft als Instrument des pädagogischen Professionalisierungsrepertoires in der Praxis zur Unterrichtsentwicklung eingesetzt werden kann, verdeutlicht Abschnitt 4.3.1. Im Abschnitt 4.3.2 wird betrachtet, welche Herausforderungen damit auf der Personal- und Schulentwicklungsebene verbunden sind.

#### **4.3.1 Ebene der Unterrichtsentwicklung**

Welche Implementierungsschritte für den Einsatz des Fragebogenverfahrens im schulischen Praxisalltag auf Lehrkraftebene mit dem langfristigen Ziel der Unterrichtsentwicklung erforderlich sind, werden in der von der Autorin konzipierten *Schülerfeedbackschleife* als Orientierungsrahmen in Abbildung 7 dargestellt und im Folgenden beschrieben. Unterrichtsentwicklung wird dabei verstanden als „...Gesamtheit der systematischen Anstrengungen..., die darauf gerichtet sind, die Unterrichtspraxis...zu optimieren. Das

grundlegende Ziel der Unterrichtsentwicklung ist die Effektivierung des Lernens der Schüler/innen in allen Dimensionen.“ (Horster & Rolff, 2001, S. 58).

Vor dem Einsatz des Schülerfeedbackverfahrens gilt es, in der *Konzeptionsphase* (1) die Intention und Voraussetzungen der Methodik des Schülerfeedbacks zu reflektieren. Das Feedbackinstrument ist als Verfahren zur Standortbestimmung bezüglich multipler Unterrichtszielkriterien konzeptualisiert und ermöglicht in Kombination mit dem Lehrerfragebogen den Abgleich mit der Selbsteinschätzung der Lehrkraft. Das Fragebogentandem fungiert dabei nicht als Beurteilungs- oder Kontrollinstrument, sondern – entsprechend dem Verständnis unterrichtlichen Geschehens als Resultat komplexer Interaktionsprozesse einer Reihe von Einflussfaktoren – als unterstützendes Hilfsinstrument zur Einschätzung von Lern- und Unterrichtsprozessen. Das Fragebogenverfahren vermittelt dabei empirisch relevante Unterrichtsqualitätsmerkmale (Weiterbildungsfunktion) und erlaubt das empirisch begründete Einleiten von Veränderungen zur Verbesserung der Lehr-Lernprozesse (Entwicklungsfunktion), indem es an den Prozessmerkmalen des Unterrichts und somit den Stellschrauben zur Outcome-Veränderung ansetzt.

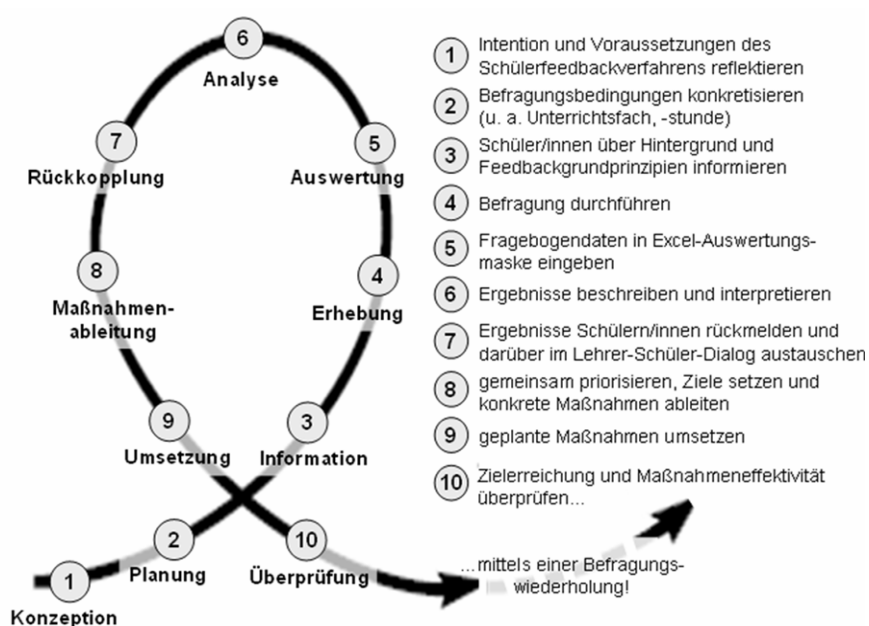


Abbildung 7: Schülerfeedbackschleife zum Einsatz des Fragebogentandems für Lehrkräfte im Rahmen der Unterrichtsentwicklung

Die *Planungsphase* (2) zielt auf die Konkretisierung der Befragungsbedingungen ab, d. h. zu welchem Zeitpunkt, in welchem Unterrichtsfach und wie genau die Erhebung erfolgen soll. Die Erhebungen zur Überprüfung des Schülerfragebogens (s. Kapitel 6 und 7 zeigten, dass ca. 15 Minuten für die parallele Bearbeitung des Lehrer- und der Schülerfragebögen



anzusetzen sind und zur Einschätzung einer Unterrichtsstunde vollen zeitlichen Umfangs die Befragung zu Beginn der zweiten Doppelstundenhälfte sehr praktikabel ist. Die Befragung sollte ohne Zeitbegrenzung erfolgen, so dass alle Schüler die Möglichkeit haben, auch bei unterschiedlicher Lese- bzw. Antwortgeschwindigkeit den gesamten Fragebogen auszufüllen. Der Fragebogen dient der Statusdiagnostik, d. h. der Erfassung des Ist-Zustandes, kann in dieser Funktion aber auch z. B. bei einer wiederholten Befragung nach der Durchführung einer Intervention durch die Lehrkraft in einem längsschnittlichen Analysezugang zur Anwendung kommen. Werden die Rahmenbedingungen weitgehend konstant gehalten und der zeitliche Befragungsabstand so gewählt, dass der Einfluss von Erinnerungseffekten möglichst gering ist, zeigt der Ergebnisvergleich, ob und welche Veränderungen die Schüler wahrnehmen. In der zur Überprüfung herangezogenen Stichprobe wird der Bezug der Schülerangaben auf eine konkrete Unterrichtsstunde gewählt. Denkbar ist auch, den Bezugsrahmen auf einen anderen Zeitraum des Unterrichtsgeschehens festzusetzen, z. B. eine Doppelstunde oder eine Unterrichtseinheit. Auch die fallweise Erhebung vor bzw. nach besonderen Anlässen, wie z. B. gruppenarbeitsintensiven Phasen, kann im Einzelfall hilfreich sein. Aufgrund des kontinuierlichen Entwicklungsfokus steht der regelmäßige Fragebogeneinsatz dabei im Vordergrund.

Die anschließende *Information* (3) der Feedbackgeber über die Intention der Unterrichtsrückmeldung ist essentieller Bestandteil eines effektiven Feedbackprozesses. Freiwilligkeit und Anonymität als Grundprinzipien der Durchführung von Schülerfeedback müssen ebenfalls deutlich werden. Die Anonymität lässt sich durch Aggregation der Daten, die keinen Rückschluss auf den Einzelnen zulässt sowie ein entsprechendes Erhebungsprozedere sicherstellen. Zusätzlich zur Zieltransparenz sollte dargelegt werden, wer Einsicht in die Daten erhält, denn Schüler beziehen potentielle Datenrezipienten in ihre Angaben ein (Elsner & Börner, 2000). Daran schließt sich die *Erhebungsphase* (4) an, die auf eine möglichst störungsfreie Fragebogenbearbeitung abzielt. Einblicknahmen Dritter sollten während der Erhebung ausgeschlossen und die Fragebögen am besten von einem Schüler ausgeteilt und eingesammelt werden.

Die statistische *Auswertung* (5) erfolgt anhand einer Excel-Auswertungsmaske, die vom Projektmitglied Wolfgang Wagner erstellt wurde und den Lehrkräften zusammen mit dem Fragebogen zur Verfügung gestellt wird. Die Datenhoheit liegt allerdings beim Feedbackempfänger. Nach Dateneingabe liefert die Auswertungsmaske numerische Basiskennwerte

sowie eine graphische Darstellung der Antwortverteilung und ein Profil, in welchem die Selbsteinschätzung der Lehrkraft dem Klassenmittelwert für die erfassten Unterrichtsmerkmale auf Itemebene gegenübergestellt wird. Die inhaltliche *Analyse* (6) der Auswertungsergebnisse sollte anschließend in zwei Schritten erfolgen: (1) In einer Phase ausschließlich deskriptiver Ergebnisbetrachtung sind die eigenen Einschätzungen im Profilverlauf nachzuvollziehen, um sich die zugrunde liegenden Verhaltenssequenzen in Erinnerung zu rufen. Darauf folgt die Betrachtung der Schülerangaben und schließlich der Abgleich dieser mit den Lehrkraftangaben hinsichtlich qualitativer und quantitativer Wahrnehmungsunterschiede auf deskriptiver Ebene. (2) Erst danach folgt die *Interpretation* der deskriptiven Befunde zu möglichen Ursachen des spezifischen Ergebnisprofils und eventuell vorhandener Zusammenhänge. Sind beispielsweise Abweichungen auf unterschiedliche Unterrichtsereignisse zurückzuführen, die hinsichtlich der Einschätzung herangezogen wurden, differieren die Erwartungen oder wird die Handlungsintention der Lehrkraft von den Schülern nicht erkannt? Hierbei gilt es in Anlehnung an die Ausführungen im Abschnitt 3.5 zu beachten, dass der fächerübergreifende Fragebogen immer nur eine Annäherung an die Komplexität des Unterrichtsgeschehens sein kann und nicht alle Items für den evaluierten Unterrichtsausschnitt einer Klasse relevant sind. Geben die Schüler z. B. an, die Lehrkraft habe sie zu Stundenbeginn nicht freundlich begrüßt (Item 22), hat dies aber bereits in einem anderen Fach in dieser Klasse getan, liegt es im Ermessen der Lehrkraft, dieses Item von weiteren Analysen auszuschließen. Liegen bereits die Ergebnisse wiederholter Befragungen in einer Klasse vor, ist die individuelle Bezugsnorm bei der Ergebnisreflexion zu präferieren. Der Vergleich mit sachlichen Standards (kriteriale Bezugsnorm) ist aufgrund der Verschränkung des Fragebogens mit dem Orientierungsrahmen Schulqualität in Rheinland-Pfalz inhärent, allerdings muss immer die spezielle Klassen- und Unterrichtssituation in die Bewertung einbezogen werden (Centra, 1993). Der soziale Vergleich mit Kollegen kann in Form kollegialer Hospitationen anhand von Beobachtungsaufträgen ebenfalls als hilfreich erlebt werden.

Die *Rückkopplung* (7) der Ergebnisse durch die Lehrkraft an die Schüler stellt einen essentiellen Schritt im Feedbackprozess dar. Die im Lehrer-Schüler-Dialog mit den Erwartungen, Interpretationen und Meinungen der Schüler angereicherte Ergebnisdarstellung verdeutlicht, dass das Schüler- und Lehrerverhalten interagiert, beide die Unterrichtsqualität bedingen und daher auch auf Schülerseite eine gewisse Selbstreflexion notwendig ist (Waibel, 2003). Der Lehrer-Schüler-Austausch bedingt und etabliert gleichzeitig eine

entsprechende Feedbackkultur, die sich als möglichst beschreibend, nicht wertend, wertschätzend und anhand von konkreten Verhaltensbeispielen nachvollziehbar charakterisieren lässt. Der Austausch mit den Schülern schwächt einige interpretative Hypothesen durch die Lehrkraft ab und verstärkt andere; dadurch können Entwicklungserfordernisse mit beiderseitig zugestandenem Handlungsbedarf erkannt werden. Diese bilden die Basis für die Formulierung des Soll-Zustandes und die Ableitung konkreter *Maßnahmen* (8). Eine gemeinsame Maßnahmenvereinbarung und schriftliche Fixierung dieser erhöht die Motivation der Schüler zur Umsetzung geplanter Interventionen. Der Rahmen denkbarer Maßnahmencharakteristika ergibt sich aus dem 4-Felder-Schema mit den Dimensionen schüler- vs. lehrerbezogen und kurz- vs. langfristig. Ein kurzfristiges schülerbezogenes Ziel ist z. B. die Visualisierung der Klassenregeln im Klassenzimmer aufgrund eines geringen Itemmittels beim Item 15 („Die Klassenregeln waren mir bekannt“). Ein eher langfristiges lehrerbezogenes Ziel ist die Etablierung einer konstruktiven Fehlerkultur als Folge geringer Zustimmungswerte bei Item 21 („Mit Fehlern ist die Lehrkraft so umgegangen, dass ich daraus etwas lernen konnte“). Helmke und Hosenfeld (2005) nennen u. a. die Erprobung verschiedener Lehr-Lernszenarien, den Austausch zu videografierten Ausschnitten des eigenen Unterrichts oder die Teilnahme an Schulungsmaßnahmen als mögliche Interventionen auf der Lehrerebene.

Die konsequente *Umsetzung* (9) geplanter Maßnahmen sichert die Nachhaltigkeit des Schülerfeedbacks, denn die Schüler nehmen das Verfahren langfristig nur dann ernst, wenn sie merken, dass ihr Feedback auch etwas bewirkt. Dabei geht es nicht nur um Optimierungsinterventionen, sondern auch darum, Bewährtes zu stabilisieren und Erreichtes bewusst zu würdigen (Schratz, 1999). Schülerfeedback rekuriert aufgrund seiner Entwicklungsfunktion auf einen kontinuierlichen und länger andauernden zyklischen Prozess der Diagnose, Intervention und Überprüfung. Umgesetzte Maßnahmen werden somit hinsichtlich ihrer Wirkungen durch erneute Befragungen überprüft. Dadurch können ergänzende Veränderungsimpulse generiert und neue Interventionen angestoßen werden. Die *Überprüfungsphase* (10) lässt die Feedbackschleife somit von Neuem beginnen – mit einem reduzierten Vorbereitungsaufwand aufgrund bereits erworbener Erfahrungen und dem zusätzlichen Fokus auf einem längsschnittlichen Daten- und Erkenntnisabgleich.

### 4.3.2 Ebene der Personal- und Organisationsentwicklung

Die Unterrichts- und Schulentwicklung sind gemäß Bastian (2007) aufeinander angewiesen, wobei die Personalentwicklung ein verbindendes Glied darstellt. Um den in Abschnitt 4.3.1 dargestellten Schülerfeedbackprozess auf Lehrkraftebene als Unterrichtsentwicklungsmaßnahme zu realisieren, sind vorbereitende, begleitende und nachbereitende Funktionen von Seiten der Schulleitung sowie pädagogischer Serviceeinrichtungen zu übernehmen.

Zur *Prozessvorbereitung* zählt die Information der Lehrkräfte über die Möglichkeit und Grundidee des Schülerfeedbacks. Die Funktion der Selbstevaluation als Instrument zur kritischen Selbstreflexion im Rahmen vorhandener schulischer Autonomiespielräume sollte dabei deutlich werden. Im Anhang sind eine Kurzbeschreibung des Fragebogenverfahrens (Anhang A.1) sowie eine kommentierte Literaturliste mit praxisnahen Ausführungen und Tipps zum Schülerfeedbackverfahren (Anhang A.2) als Begleitmaterialien für die Phase der Bekanntmachung des Verfahrens zu finden. Es ist wichtig, dass die Schulleitung hinter dem Schülerfeedbackansatz steht und das Verfahren im eigenen Unterricht anwendet. Den im Abschnitt 4.1 aufgeführten Charakteristika der Selbstevaluation Rechnung zu tragen, ist aufgrund der notwendigen Selbststeuerung die Freiwilligkeit der Teilnahme der Lehrkräfte oberstes Gebot. Die Hoffnung, dass sich *alle* Lehrkräfte für die Idee begeistern, ist nicht realistisch; die Anwendung durch eine ‚kritische Masse‘, welche eine Kulturveränderung anstößt, durchaus plausibel. Moderatoren<sup>7</sup> können diesbezüglich eine Vorreiter- und Informationsrolle übernehmen. Lehrkräfte, die das Verfahren bereits in der Überprüfungsphase des Fragebogens eingesetzt haben, könnten als Multiplikatoren oder Mentoren gewonnen werden (s. Kapitel 8). Ferner müssen das fachlich notwendige Know-How zur Evaluationsdurchführung bzw. unterstützende externe Expertise durch die pädagogischen Unterstützungssysteme zielorientiert vermittelt werden, um eine Überforderung auch im Hinblick auf die hohe Belastung aufgrund der Vielfalt an Rollenansprüchen, denen die Lehrkräfte im Evaluationsprozess gerecht werden müssen, zu vermeiden.

Während des Schülerfeedbackprozesses ist es wichtig, die notwendigen Ressourcen und Unterstützungsangebote zur Verfügung zu stellen, ohne die Lehrkraft in ihrer Selbststeuerung einzuschränken. Besonders in der Phase der Maßnahmenableitung und

---

<sup>7</sup> (Fach- und Schulentwicklungs-)Moderatoren sind in Rheinland-Pfalz an eine pädagogische Serviceeinrichtung angebunden. Sie sind hinsichtlich fachlicher und prozessbezogener Kompetenzen qualifiziert und unterstützen Schulen in der Qualitätsentwicklungsarbeit (Arbeitsgruppe Qualitätsstandards, 2007).

-umsetzung müssen alle pädagogischen Unterstützungssysteme versuchen, einen an den Bedürfnissen der Lehrkraft orientierten Entwicklungssupport zu leisten. Im Schülerfeedbackprozess entstehende Unsicherheiten auf Seiten der Lehrkräfte müssen ernst genommen und im Austausch über Chancen und Grenzen des Schülerfeedbacks aufgefangen werden. Hierzu bieten sich kollegiale Fallberatungen an. Je nach personellen Voraussetzungen und Bedarf könnten diese Dialoggruppen fachbezogen oder fachübergreifend zusammengestellt werden und sich regelmäßig nach einer vorgegeben Gesprächsstruktur wechselseitig zu Fragen und Schlüsselthemen des Schülerfeedbacks beraten bzw. gemeinsam mögliche Lösungen entwickeln.

*Nach* Abschluss der ersten Schülerfeedbackschleife sollte die Lehrkraft neben der Möglichkeit zur Reflexion mit Kollegen auch erkannte Defizite durch entsprechende Qualifizierungsmaßnahmen ausgleichen können, z. B. mit Hilfe entsprechender Angebote der pädagogischen Unterstützungssysteme. Das Schülerfeedback ist dabei nicht als isolierter Ansatz zu verstehen, sondern sinnvoll durch weitere Perspektiven zu ergänzen, welche die Anzahl und Vielfalt stimulierender Unterrichtsentwicklungsimpulse erhöhen. Potentielle Feedbackquellen sind in Abbildung 8 im Feedbackrad aufgeführt: Denkbar sind neben dem bereits erwähnten Rückgriff auf die Wahrnehmungen der (Fach-)Kollegen im Rahmen kollegialer Fallberatungen bzw. gegenseitiger Unterrichtshospitationen mit anschließendem Austausch, die mündliche oder schriftliche Befragung der Eltern, externe Beobachterrückmeldungen im Anschluss an Unterrichtsbesuche sowie die Einschätzung videografierten Unterrichts. Auch der Vergleich mit Referenzwerten kongruenter Befragungen bei Übereinstimmung der Schulart, der Klassenstufe sowie des Unterrichtsfachs und optimalerweise bei vergleichbarem Unterrichtsinhalt sowie ähnlichen regionalen Kontextbedingungen ist sinnvoll.

Um den langfristigen Transfer der Erkenntnisse von der Unterrichtsentwicklungsebene der einzelnen Lehrkraft auf die Ebenen der Personal- und Organisationsentwicklung der gesamten Schule im Sinne einer ‚lernenden Organisation‘ nach Rosenstiel (1999) sicherzustellen, könnten z. B. die (Fach-)Moderatoren wichtige Schlussfolgerungen zwischen den kollegialen Beratungsgruppen – unter Beachtung der Datenhoheit des Feedbacknehmers – austauschen und anhand von Vorträgen oder in Berichtsform dem gesamten Kollegium zugänglich machen. Die Integration der Erkenntnisse in einen handlungsorientierten Schulentwicklungsprozess entspricht dabei in Anlehnung an die Erfahrungen zur Mitarbeiterbefragung im organisationspsychologischen Bereich dem größtmöglichen Feedbackprozessumfang,

sichert allerdings die Nachhaltigkeit des Verfahrens (Ladwig & Domsch, 1997). Dieser kann durch einen schulübergreifenden Austausch auf Konferenzen durch die Moderatoren oder die Schulleiter auf Basis aggregierter Erkenntnisse z. B. mit Kollegen der gleichen Schulart sowie einen Austausch mit den Eltern ergänzt werden.

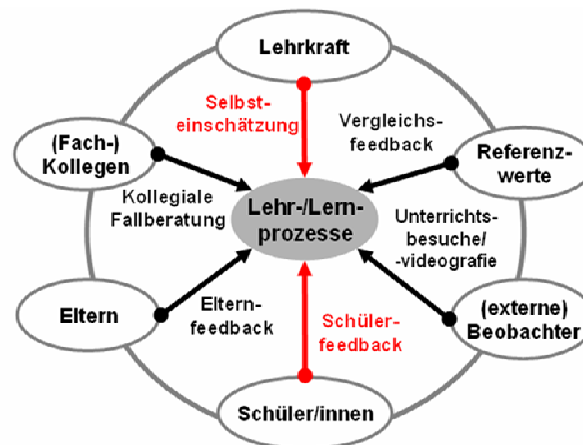


Abbildung 8: Feedbackrad sich gegenseitig ergänzender Feedbackquellen für das komplexe Unterrichtsgeschehen, wobei das Fragebogentandem des Projekts die Selbst- und Fremdeinschätzung durch die Schüler abdeckt

## 5 Präzisierung der Zielsetzungen

Das Hauptziel der vorliegenden praxisbezogenen Arbeit ist die Konzeption und Erprobung eines Fragebogens zur Erfassung der Unterrichtsqualität im Rahmen einer Selbstevaluation. Dieser Prozess kann in unterschiedliche Subziele aufgeteilt werden, die im Folgenden skizziert werden und in Tabelle 5 überblicksartig dargestellt sind.

Das erste Ziel besteht in der Überführung des Itempools in eine Fragebogenfassung für die verschiedenen Versionen des Schüler- und Lehrerfragebogens mit ich- vs. wir-Bezug und weiblicher vs. männlicher Lehrkraftbezeichnung. (2) Aufgrund der im Abschnitt 3.4 dargestellten Befunde zur Problematik der Itemformulierung soll anschließend eine Pilotierung des Itempools anhand von kognitiven Pretestinterviews durchgeführt und die Fragebogen-versionen entsprechend der Erkenntnisse modifiziert werden. (3) Der nächste Untersuchungsschritt ist die Erprobung des Fragebogens in der Sekundarstufe I. Hierbei wird eine Stichprobengröße von ca. 40-60 Klassen bei einer Gleichverteilung der Stichprobenelemente hinsichtlich der Schulart, der Klassenstufe, der erhobenen Kernfächer sowie der Geschlecht der Lehrkraft angestrebt. (4) Im Anschluss an die Datenerhebung sollen die Ergebnisse der Lehrkraft zurückgemeldet werden. Hierbei ist der Rückgriff auf die im

Abschnitt 4.2 beschriebenen Aspekte einer rezeptions- und reflexionsdienlichen Ergebnismeldung erforderlich. (5) Die nächste Phase umfasst die testtheoretische Überprüfung der Items zur Unterrichtswahrnehmung sowie die Untersuchung der faktoriellen Itemvalidität. Dabei sollten die Schülerwahrnehmungen einer deskriptivstatistischen Auswertung sowie Item- und Reliabilitätsanalysen unterzogen und anhand eines explorativen sowie konfirmatorischen faktorenanalytischen Designs hinsichtlich der Itemeignung untersucht werden. In Bezug auf die Itemselektion soll eine Balance zwischen dem wissenschaftlichen Konstruktionsanspruch und praktischen Erfordernissen in Form des spezifischen Anwendungskontexts in der Unterrichtsentwicklung erzielt werden. Die Beachtung methodisch-wissenschaftlicher und praktischer Erfordernisse ist generell ein wichtiges Kriterium für das Vorgehen in dieser Arbeit.

Die Auswertungen sollten dabei, dem späteren Anwendungskontext des Fragebogens entsprechend, auf Itemebene erfolgen. Denn bei der Unterrichtsentwicklung liegt der Fokus auf der konkreten Verhaltensebene, die Angabe von Skalenmittelwerten oder eines Gesamtwerts des Fragebogens sind nicht indiziert. Dimensionsanalytische Untersuchungen zur Faktorenstruktur werden somit nur als ergänzende Information in Bezug auf die Itemgüte hinsichtlich der Erfassung der intendierten Konstruktfacette herangezogen. Ferner erhebt der Fragebogen nicht den Anspruch, unabhängige Faktoren zu erfassen.

Aus dem verwendeten Fragebogendesign lassen sich die folgenden Fragestellungen dieser Arbeit ableiten: (5.1) Die Überprüfung der Eignung des Fragebogens hinsichtlich seiner unterschiedlichen Versionen des ich- vs. wir-Bezugs sowie (5.2) die Exploration zu dem Aspekt, inwiefern die Erfassung der Urteilsicherheit im Rahmen von Unterrichtseinschätzungen einen Informationsgewinn in Bezug auf die stattfindenden Urteilsprozesse erlaubt. Die erstgenannte Fragestellung soll anhand eines Vergleichs der Ergebnisse testtheoretischer und dimensionsanalytischer Untersuchungen beurteilt werden. Die Angaben zur Urteilsicherheit sollen deskriptivstatistisch analysiert werden und hinsichtlich der Möglichkeit geprüft werden, als Gewichtungsfaktor für die Unterrichtseinschätzungen zu fungieren. (6) Aufgrund der praxisbezogenen Ausrichtung der Arbeit stellen auch die Entwicklung und die Bereitstellung von Hilfsmaterialien für Lehrkräfte hinsichtlich des Einsatzes des Schülerfeedbackverfahrens als übergeordnetes Konstruktionsziel des Fragebogens eine Zielsetzung dieser Arbeit dar. Die entsprechenden Ausführungen dazu wurden bereits im letzten Kapitel des Theorieteils dargelegt.

Tabelle 5: Überblick zu den Zielsetzungen der Arbeit

Aspekt	Zielsetzung	Vorgehen	Ab-schnitt
(1) Formale Fragebogen-erstellung	<ul style="list-style-type: none"> <li>Erstellen der unterschiedlichen Fragebogenversionen</li> </ul>	<ul style="list-style-type: none"> <li>Layouterstellung</li> <li>Instruktionsverfassung</li> </ul>	
(2) Pilotierung der SFB-/LFB-Vorversion	<ul style="list-style-type: none"> <li>Prävalidierung des Itempools</li> </ul>	<ul style="list-style-type: none"> <li>Durchführung kognitiver Pretestinterviews</li> </ul>	6.1-6.2
(3) Erprobung der SFB-Überprüfungsversion	<ul style="list-style-type: none"> <li>Rekrutierung der Lehrkraftstichprobe (Gleichverteilung hinsichtlich Schulart, Klassenstufe, Kernfächer und Geschlecht angestrebt)</li> <li>Fragebogenerhebung unter Berücksichtigung des Untersuchungsdesigns</li> </ul>	<ul style="list-style-type: none"> <li>Durchführung eines mehrstufigen Akquiseverfahrens (in Zusammenarbeit mit IFB)</li> <li>Instruktion der Lehrkräfte</li> <li>Versand der Klassensets</li> <li>Durchführung aller Befragungen im Großraum Mannheim/Ludwigshafen (bei jeder Schulart)</li> </ul>	7.1
(4) Ergebnismeldung	<ul style="list-style-type: none"> <li>Rückmeldung deskriptiver Ergebnisse an die befragten Lehrkräfte</li> </ul>	<ul style="list-style-type: none"> <li>Verfassen lehreradäquater Kennwertbeschreibungen</li> <li>Profilardarstellung von Wolfgang Wagner umgesetzt</li> </ul>	7.4
(5) Itemüberprüfung	<ul style="list-style-type: none"> <li>Überprüfung der Items zur Unterrichtswahrnehmung hinsichtlich ihrer Konstrukt-, Inhalts- und Augenscheinvalidität</li> </ul>	<ul style="list-style-type: none"> <li>testtheoretische Itemüberprüfung</li> <li>dimensionsanalytische Untersuchungen mittels faktorenanalytischer Verfahren</li> </ul>	9.1-9.4
(5.1) ich- vs. wir-Version	<ul style="list-style-type: none"> <li>ich- vs. wir-bezogen formulierte Items hinsichtlich ihrer testtheoretischen Eignung und Anwendbarkeit im Fragebogenkontext vergleichen</li> </ul>	<ul style="list-style-type: none"> <li>Vergleich der Analyseergebnisse der testtheoretischen und dimensionsanalytischen Untersuchungen, s. (5)</li> </ul>	11.1
(5.2) Analyse der Urteilssicherheit	<ul style="list-style-type: none"> <li>Urteilssicherheitsangaben der Schüler hinsichtlich ihres Informationsgehalts prüfen</li> </ul>	<ul style="list-style-type: none"> <li>deskriptivstatistische Untersuchungen</li> <li>Option als Gewichtungsfaktor der Unterrichtseinschätzung prüfen</li> </ul>	10
(6) Einsatz des Fragebogenverfahrens in der Praxis	<ul style="list-style-type: none"> <li>Bereitstellung von Hilfsmaterialien für Lehrkräfte hinsichtlich des Einsatzes des Schülerfeedbackverfahrens</li> </ul>	<ul style="list-style-type: none"> <li>Ausarbeitung von Informationsmaterial</li> <li>Entwicklung eines Rahmenmodells zum Einsatz von Schülerfeedback</li> </ul>	4.3.1

Anmerkung: SFB = Schülerfragebogen, LFB = Lehrerfragebogen



## II Empirischer Teil

Im empirischen Teil werden die erforderlichen Schritte zur Weiterentwicklung und Überprüfung des Itempools zum Erfassen der Unterrichtsqualität dargestellt: Nach der Beschreibung des Prozedere der qualitativen Voruntersuchung (Kapitel 6) und der Haupterhebung (Kapitel 7) folgen wesentliche Aspekte der Analysevorbereitung (Kapitel 8). Danach werden zuerst die Items zur Unterrichtswahrnehmung (Kapitel 9) und anschließend die Urteils sicherheitsangaben hinsichtlich ihrer Eignung untersucht (Kapitel 10). Kapitel 11 integriert die Analyseergebnisse in Bezug auf die endgültige Itemauswahl. Bevor die Analyseergebnisse sowie weiterführende Aspekte in Kapitel 12 diskutiert werden.

### 6 Qualitative Voruntersuchung

Nach Erstellung der SFB-Vorversion (s. Anhang B.1) wurde diese im Oktober 2007 mit Hilfe kognitiver Pretestinterviews überprüft. Im folgenden Abschnitt werden nach einer allgemeinen Beschreibung der Vorgehensweise (Abschnitt 6.1) wesentliche Ergebnisse der qualitativen Voruntersuchung aufgeführt (Abschnitt 6.2).

#### 6.1 Durchführung kognitiver Pretestinterviews

Als qualitative Voruntersuchungsmethode wird in Anlehnung an Kurz, Prüfer und Rexroth (1999) das kognitive Pretestinterview mit dem Ziel eingesetzt, aufschlussreiche Hinweise auf das Itemverständnis, das den Antworten zugrunde liegt, zu erhalten. Dabei liegt der Fokus allerdings weniger auf der detaillierten Abbildung der kognitiven Prozesse, die zu den Antworten führen, sondern vielmehr darauf, Verständnis- und somit Validitätsprobleme von Items aufzudecken. Der Hintergrund ist, dass ein komplizierter Satzbau, vage Begriffe oder Fremdwörter sowie eine nicht alltägliche Perspektive Verständnisprobleme hervorrufen könnten, u. a. bei Schülern mit Migrationshintergrund. Zum Erfassen des Itemverständnisses werden in Anlehnung an Kurz et al. (1999) folgende *Interviewmethoden* eingesetzt:

- Paraphrasing mit der Standardinstruktion: „Bitte wiederhole die Frage mit deinen eigenen Worten!“

- Thinking aloud mit der Standardinstruktion: „Als du eben die Frage beantwortet hast, an was hast du da gedacht? Was ist dir alles durch den Kopf gegangen, bevor du die Antwort gegeben hast?“
- Probing mit der Standardinstruktion: „Stell dir vor, du müsstest jemandem den Begriff Arbeitsplatzgestaltung (Item SA61), Zeitplanung und -einteilung (Item SA66) etc. erklären, was würdest du sagen?“
- Confidence Rating mit der Standardinstruktion: „Wie sicher warst du dir bei deiner Antwort?“

Des Weiteren wird in den Pretestinterviews die benötigte Bearbeitungsdauer für die SFB-Vorversion erfasst, die Bearbeitungsinstruktion und die Antwortkategorien hinsichtlich ihres Verständnisses überprüft und spezifische Reihenfolge- bzw. Kontexteffekte ermittelt. Bei letzteren wird durch das Abfragen zweier Items in umgekehrter Reihenfolge untersucht, inwiefern die Reihenfolge eine Auswirkung auf das Verständnis der beiden Items hat. Ferner wird nach bisherigen Interviewerfahrungen, dem Teilnahmegrund und nach der persönlichen Bewertung des Interviews gefragt. Hintergrund ist, mögliche Verzerrungen aufgrund von zu lang empfundenen Interviews, geringer Motivation oder Interviewerfahrung in Betracht ziehen zu können.

Die Probanden wurden in Mannheim rekrutiert und ca. 40-60 Minuten anhand eines standardisierten Interviewleitfadens in der Stadtbücherei Mannheim mit einer Vergütung von 5,- EUR befragt (s. Anhang B.2). Die Lese- und Bearbeitungszeiten für die Instruktion, die Erhebung soziodemographischer (Geschlecht, Klassenstufe, Schularart und Sprachkenntnisse) sowie allgemeiner Daten (Unterrichtsfach, Noten) und für die Beantwortung der fünf Unterrichtsmerkmalsskalen wurden erfasst. Zur Untersuchung, welche Form der Lehrkraft-Bezeichnung bevorzugt wird, wurde den Probanden jeweils der gleiche Ausschnitt der SFB-Vorversion in drei verschiedenen Versionen in unterschiedlicher Reihenfolge (s. Anhang B.3) vorgelegt. Die Probanden wurden anschließend nach ihrer Präferenz bezüglich der drei Formulierungen „die Lehrkraft“ vs. „die Lehrerin/der Lehrer“ vs. „die Lehrperson“ gefragt. Die Fragebogenverbesserung erfolgte sukzessive, d. h. zeigten sich bei den ersten fünf Probanden die gleichen Verständnisschwierigkeiten bei einem Item, so wurde dieses überarbeitet und weiteren Probanden zur Überprüfung vorgelegt.

Die in der Voruntersuchung befragte *Stichprobe* soll repräsentativ für die Stichprobe der Haupterhebung sein. Deswegen wurde darauf geachtet, dass diese möglichst das gesamte

Spektrum der jeweiligen Kategorien der Merkmale Geschlecht, Klassenstufe (7. bis 10. Klasse), Schulart (Haupt-, Realschule, Gymnasium) sowie Sprachkenntnisse in Deutsch (gering, mittel, Muttersprache) abdecken. Hinsichtlich des beurteilten Unterrichtsfachs sollten möglichst alle Kernfächer (Mathematik, Deutsch, Englisch, Französisch, Physik, Chemie, Biologie) vertreten sein. Die untersuchte Stichprobe ( $n = 14$ ) setzt sich aus 6 Mädchen und 8 Jungen zusammen und weist hinsichtlich der Merkmale Schulart, Klassenstufe und Deutschkenntnisse die in Tabelle 6 dargestellte Verteilung der Stichprobenelemente auf. Die Notendurchschnitt beträgt 3.11 über alle erfassten Fächer und 3.07 im beurteilten Unterrichtsfach, wobei die Fächer mit folgenden Häufigkeiten vertreten sind:  $n_{\text{Mathematik}} = 3$ ,  $n_{\text{Deutsch}} = 4$ ,  $n_{\text{Englisch}} = 1$ ,  $n_{\text{Französisch}} = 1$ ,  $n_{\text{Physik}} = 2$ ,  $n_{\text{Chemie}} = 2$  und  $n_{\text{Gemeinschaftskunde}} = 1$ .

Darüber hinaus wurde die Vorversion des zum SFB inhaltsgleichen *Lehrerfragebogens* von einer Realschul- und einer Gymnasiallehrkraft mit Bezug auf eine konkrete Unterrichtsstunde bearbeitet.

Tabelle 6: Verteilung der Stichprobenelemente der qualitativen Voruntersuchung hinsichtlich der besuchten Klassenstufe, Schulart und der Deutschkenntnisse

Klassen- stufe Schulart	7. Klasse	8. Klasse	9. Klasse	10. Klasse	n
Hauptschule	1 (mittel, Türkisch)	1 (mittel, Bulgarisch)	2 (gering, Albanisch; mittel, Türkisch)		4
Realschule		2 (Muttersprache; mittel, Berbisch)	3 (Muttersprache; mittel, Kurdisch, mittel, Spanisch)	1 (Muttersprache)	6
Gymnasium			1 (Muttersprache)	3 (Muttersprache; Muttersprache; Muttersprache)	4
n	1	3	6	4	14

**Anmerkungen:** Der Wert in Klammern innerhalb der Zellen gibt die Deutschkenntnisse in der Abstufung gering, mittel und Muttersprache an. Bei mittleren Deutschkenntnissen folgt die Angabe der Sprache, die neben Deutsch *auch* zu Hause gesprochen wird; bei geringen Deutschkenntnissen ist die Sprache angegeben, die *meistens* zu Hause gesprochen wird.

## 6.2 Darstellung der Interviewergebnisse

Die qualitative Voruntersuchung des *Schülerfragebogens* erbrachte folgende Ergebnisse:

- Die durchschnittliche Bearbeitungsdauer liegt bei 13.5 Minuten mit einem Streubereich von [9 min 17 sec; 20 min 30 sec].
- Die Bearbeitungsinstruktion wird von allen Probanden korrekt verstanden.
- Die Antwortkategorien der Unterrichtswahrnehmung fassen 11 von 14 Probanden symmetrisch auf - sie geben hinsichtlich der Beschreibung der Urteilsabstufungen mit eigenen Worten, z. B. bei „stimme nicht zu“ an, dass es für sie „ich denke nicht genauso“ oder „absolut anderer Meinung sein“ bedeutet, bei „stimme eher nicht zu“ = „tendiere dazu, nicht so zu denken“ oder „nicht sicher, eher negative Meinung“, bei „stimme eher zu“ = „tendiere dazu, so zu denken“ oder „genauso wie das zweite nur ins Positive“ und bei „stimme zu“ = „denke genauso“ oder „bin absolut der gleichen Meinung“.
- Die Antwortkategorien der subjektiven Urteilssicherheit werden ebenfalls von 11 Probanden als symmetrisch beschrieben und zwar folgendermaßen:
  - „sehr unsicher“ = „sicher, dass es nicht so ist“ oder
  - „zu 25% sicher“, „eher unsicher“ = „50:50 zum Negativen“ oder
  - „zu 50% sicher“, „eher sicher“ = „50:50 zum Positiven“ oder
  - „zu 75% sicher“, „sehr sicher“ = „100% sicher“
- Es zeigt sich eine klare Präferenz der Bezeichnung der Lehrkraft in der Singular-Form „Die Lehrerin ...“ bzw. „Der Lehrer ...“.
- Hinsichtlich der bisherigen Interviewerfahrungen ergibt sich, dass fünf Probanden keine und neun Probanden aufgrund der Teilnahme an Befragungen in öffentlichen Verkehrsmitteln, in der Schule, in Marktforschungsinstituten oder durch Internetbefragungen Interviewerfahrung haben.
- Die Probanden geben folgende Gründe für ihre Interviewteilnahme an: Aus Mangel an sonstigen Aktivitätsalternativen ( $H_n=4$ ), weil es wichtig ist, dass auch Schüler zu ihrer Meinung zum Unterricht befragt werden ( $H_n=3$ ), aus Interesse und Neugier ( $H_n=3$ ) und aufgrund des finanziellen Anreizes ( $H_n=3$ ) sowie aufgrund der Teilnahme eines Freundes ( $H_n=1$ ).

- Die Teilnehmer finden das Interview okay ( $H_n = 7$ ), interessant ( $H_n = 4$ ), gut ( $H_n = 2$ ) sowie spannend ( $H_n = 1$ ) und hinsichtlich der Interviewdauer angemessen ( $H_n = 11$ ) bzw. etwas zu lang ( $H_n = 3$ ).

Bezüglich des Frageverständnisses wird der Großteil der Fragen gemäß der Konstruktionsintention verstanden.

Tabelle 7 gibt einen ausschnitthaften Überblick darüber, welche Validitätsprobleme bei der SFB-Vorversion auftreten und wie die Items aufgrund dessen umformuliert bzw. ob sie aus dem Fragebogen entfernt werden.

Für das Ausfüllen des *Lehrerfragebogens* werden 15 Minuten bzw. 18 Minuten ( $M = 16,5$  Minuten) benötigt. Die Lehrkräfte merken weitere Punkte hinsichtlich der Augenscheinvalidität an: Das Item 15 („Die Klassenregeln waren allen bekannt“) erfasst einen Aspekt, der mit dem Klassenlehrer erarbeitet wird und somit bei der Beurteilung des Fachlehrers weniger relevant ist. Item 50 („Zu Beginn der Stunde habe ich Stoff aus vergangenen Stunden kurz wiederholt“) und Item 57 („Ich habe am Ende der Stunde die wichtigsten Ergebnisse zusammengefasst“) widersprechen der Idee, dass die Schüler den Unterrichtsstoff aus lerntheoretischen Gesichtspunkten selbst wiederholen bzw. zusammenfassen sollten. Ferner weisen sie darauf hin, dass sich der Großteil der Aktivierungitems auf Unterrichtsaspekte bezieht, die nur über mehrere Unterrichtsstunden realisierbar sind, z. B. Item 66 („Ich habe den Schülern/innen Hinweise zur Zeitplanung und Zeiteinteilung beim Lernen gegeben“). Diese Angaben werden aufgrund der geringen Stichprobengröße bei der Erstellung der endgültigen Fragebogenversion nicht berücksichtigt, fließen aber in die Hinweise bezüglich der Handhabung des Fragebogens ein (s. Abschnitt 4.3).

Tabelle 7: Übersicht zu Items mit Validitätsproblemen in der SFB-Pretestversion sowie deren Umformulierung bzw. Elimination hinsichtlich der SFB-Überprüfungsversion

Item der SFB-Vorversion	Validitätsproblem	Umformulierung
1. „Die Klassenregeln waren allen bekannt“ 2. ... 3. „Allen Schülern war bekannt, wie man sich im Unterricht verhalten soll“ „Der Lehrer / die Lehrerin hat so gesprochen, dass ich sie sehr gut verstehen konnte“	Sequenzeffekt derart, dass das dritte Item missverstanden wird, als <i>ob</i> sich die Schüler so verhalten haben.  „Hat so gesprochen“ wird z. T. als „inhaltlich verstanden“ interpretiert.	Nur das erste Item (entspricht SKF15) wird aufgenommen.  SKV46: „Der Lehrer vs. die Lehrerin hat so laut und deutlich gesprochen, dass ich alles sehr gut verstehen konnte“
„Wenn die Arbeitsform geändert wurde, wusste jeder gleich, was er/sie zu tun hat“	„Wenn die Arbeitsform geändert wurde...“ wird missverstanden als „wenn das Thema gewechselt wird“.	Dieses Item wird entfernt.
1. „Zu Unterrichtsbeginn hat der Lehrer / die Lehrerin erklärt, was wir in dieser Stunde lernen sollen“ 2. ... 3. „Zu Unterrichtsbeginn hat der Lehrer / die Lehrerin darauf hingewiesen, was das Ziel ist“	Sequenzeffekt derart, dass das dritte Item so verstanden wird, dass der gesamten Klasse klar ist, was in der Stunde gemacht wird, da im ersten Item bereits erfasst, dass die Lehrerin erklärt, was in der Stunde behandelt wird bzw. was das Ziel ist.	Das erste Item (entspricht SKV45) wird durch Füllitems von dem dritten Item (entspricht SKV52) deutlich getrennt.
„Wenn andere Schüler gesprochen haben, konnte ich sie sehr gut verstehen“	Wird missverstanden als: „Andere Schüler erklären den Unterrichtsinhalt besser als die Lehrkraft“	SKV54: „Die Mitschüler/innen haben so laut und deutlich gesprochen, dass ich sie gut verstehen konnte“
„In Kleingruppen haben Schüler/innen die Rolle des Lehrers (z. B. als Tutor) übernommen“	Dieses Item verstehen zwei Probanden korrekt, die anderen nicht oder in der Form, dass jeweils Kleingruppen den Lehrer spielen. Der Begriff „Tutor“ ist in 5 von 6 Fällen unbekannt.	Dieses Item wird entfernt.
„Die Schüler/innen haben den Lehrer / die Lehrerin durch Bedienung von Geräten unterstützt“	Dieses Item wird zweimal korrekt, zweimal nicht verstanden und in zwei Fällen als „Der Lehrer erklärt etwas, was der Schüler dann selbst anhand eines Beispiels an der Tafel erklären soll“ und als „Hinter der Tafel befindet sich ein Lösungsblatt und nach Aufgabenende soll man selbst vorgehen und die richtige Lösung abgleichen“ interpretiert.	SA72 „Ich vs. die Schüler/innen habe(n) den Lehrer vs. die Lehrerin durch Bedienung von technischen Geräten unterstützt“
„In der Unterrichtsstunde stellten wir uns auch gegenseitig Fragen zur Unterrichtsthematik“	„Unterrichtsthematik“ wird in der Regel nicht verstanden.	SA73: „Ich habe anderen Schülern/innen Fragen zum Unterrichtsstoff gestellt“
„Alle Schüler/innen haben sich während der Unterrichtsstunde mindestens einmal gemeldet.“	„Gemeldet“ wird durchgängig als „melden und antworten“ missverstanden.	SA78: „Alle Schüler/innen sind während der Unterrichtsstunde mindestens einmal drangekommen“

## 7 Haupterhebung

Nach der Einarbeitung der Pretestergebnisse wurden die vier verschiedenen SFB-Überprüfungsversionen mit Hilfe der Software TELEform<sup>®</sup> Enterprise (Firma Cardiff Inc., Version 8.23)<sup>8</sup> erstellt, die wir-Version mit der männlichen Lehrkraftbezeichnung kann im Anhang C.1.6 eingesehen werden. Danach begann die Durchführung der Datenerhebung, die im Abschnitt 7.1 dargestellt wird. Ein Überblick über die Datenaufbereitung ist Gegenstand des Abschnitts 7.2. In den darauf folgenden Abschnitten wird die Schüler- und Lehrkraftstichprobe beschrieben (Abschnitt 7.3) sowie das Prozedere der Ergebnisrückmeldung an die Lehrkraft skizziert (Abschnitt 7.4).

### 7.1 Durchführung der Befragung

Zur Befragungsdurchführung wurde im ersten Schritt eine anfallende Stichprobe rekrutiert, bei der eine Repräsentativität für die Kernfächer Deutsch, Mathematik, Englisch, Französisch und Naturwissenschaften der Klassenstufen 7 bis 10 der Hauptschule, der Realschule, des Gymnasiums, der Regionalen Schule, der Integrierten Gesamtschule und der Dualen Oberschule in Rheinland-Pfalz sowie das Lehrkraftgeschlecht angestrebt wurde. Lediglich die Berufsbildenden Schulen und Förderschulen wurden nicht erhoben.

Die *Akquise* erfolgte in einem mehrstufigen Vorgehen (s. Anhang C.1.1) und fokussierte die Gruppe der Moderatoren als potentielle Multiplikatoren des Schülerfeedbackverfahrens (s. Abschnitt 4.3.2) sowie in Abhängigkeit vom Rücklauf weitere Lehrkräfte auch ohne eine solche Zusatzqualifizierung. Rekrutierungsmittel waren eine Projektausschreibung auf der Internetseite des IFB (s. Anhang C.1.2) und die Ansprache sowohl der Verantwortlichen für die Moderatorengruppen als auch potentieller Lehrkräfte und Schulleiter. Das Vorhaben wurde auch in einschlägigen Fortbildungsveranstaltungen des IFB vorgestellt. Die gezielte Ansprache von Lehrkräften und Schulleitern fand in der zweiten Akquisestufe statt, dabei wurde verstärkt auf die angestrebte Gleichverteilung der Stichprobenkriterien geachtet. Mit allen Lehrkräften, die ihre Bereitschaft zur Teilnahme erklärt hatten, wurden der Projektablauf und die *Befragungsdurchführung* telefonisch besprochen. Die Befragung fand dabei direkt im Anschluss an die zu beurteilende Unterrichtsstunde in einer von der Lehrkraft selbst ausgewählten Klasse statt – entweder in einer Doppelstunde oder in Ab-

sprache mit den Kollegen in der nachfolgenden Unterrichtsstunde. Die Bearbeitungsdauer der Fragebögen betrug unabhängig von der Schulart inklusive der Instruktion zwischen 15-25 Minuten. Die Fragebögen waren hinsichtlich der Urteilsperspektive (ich- vs. wir-Bezug) vorsortiert und wurden von der Lehrkraft so ausgeteilt, dass nebeneinander sitzende Schüler jeweils unterschiedliche Versionen erhielten. Die Lehrkraft bearbeitete zeitgleich den zum Schülerfragebogen äquivalenten Lehrerfragebogen.

Im Regelfall führte die Lehrkraft die Befragung selbst in ihrer Klasse durch. Da es sich bei der Befragung allerdings um die Erprobung eines neuen Fragebogenverfahrens handelt, befragte die Autorin jeweils bei einer Schule pro Schulart mindestens eine Klasse selbst vor Ort. Dabei wurden sowohl die Klasse als auch die Lehrkraft im Anschluss an die Fragebogenbearbeitung in Anlehnung an die Verfahrensweise bei den Pretestinterviews hinsichtlich eventueller Probleme bei der Fragebogenbearbeitung befragt, um diese präventiv in die Instruktion für die darauf folgenden Befragungen aufnehmen zu können. Führte die Lehrkraft die Befragung in ihrer Klasse selbst durch, klärte die Autorin die Rahmenbedingungen (Zeitpunkt, Ablauf, etc.) telefonisch bzw. per E-Mail ab und verschickte die Fragebögen mit der entsprechenden Lehrkraftbezeichnung (die Lehrerin vs. der Lehrer) zusammen mit einem Anschreiben (s. Anhang C.1.3), einer Standardinstruktion (s. Anhang C.1.4) und einem frankierten Rückumschlag. Dem Klassensatz lag auch die Bitte bei, dass die Schüler ihre letzte (Halbjahres- bzw. Endjahres-)Note im eingeschätzten Fach auf dem Fragebogen vermerken sollten. Ferner sollten die Lehrkräfte die Schulleitung bezüglich der Befragung anhand eines Projektschreibens (s. Anhang C.1.5) informieren, sofern dies nicht bereits ohnehin durch die Autorin erfolgt war.

Der *Erhebungszeitraum* wurde in der Planungsphase von Anfang November bis Ende Januar 2008 angesetzt, musste allerdings bis März 2008 verlängert werden. Gründe dafür waren, dass das Schuljahr 2007/2008 sehr kurz war, im November sowie Dezember 2007 konzentriert Klassenarbeiten geschrieben wurden, bereits im Januar Zeugniskonferenzen stattfanden, Besuche der AQS-Evaluationsteams im Rahmen der externen Evaluation und die Umsetzung von Strukturreformmaßnahmen (v. a. bei den Hauptschulen) zusätzliche Zeitressourcen forderten und die Befragungen seitens der Lehrkräfte häufig nicht termingerecht durchgeführt werden konnten. Nach Abschluss der Erhebung lagen schließlich 1741

---

<sup>8</sup> TELEform<sup>®</sup> ist eine Software zur Verarbeitung von Formularen und ermöglicht u. a. die elektronische Fragebogenerstellung (TELEform<sup>®</sup>-Designer) und Datenerfassung (TELEform<sup>®</sup>-Reader).



Schüler- und 87 Lehrerfragebögen vor. Im folgenden Abschnitt wird beschrieben, wie diese Daten für die geplanten Analysen aufbereitet wurden.

## 7.2 Aufbereitung der Rohdaten

In einem ersten Schritt wurden alle Fragebögen einzeln hinsichtlich offener Anmerkungen oder fehlender bzw. systematischer Antworten vorkontrolliert. Entsprechende Auffälligkeiten wurden im Hinblick auf eine Verwertung bei der Missinganalyse vermerkt. Danach erfolgten das Scannen der Fragebögen und die Überführung in ein für Statistikprogramme lesbares Format. Die im Schüler- und Lehrerdatsatz erhobenen Variablen können den entsprechenden Codebooks sowie der Variablenübersicht (s. CD-Rom) entnommen werden. Beide Datensätze wurden anschließend ausführlichen fallweisen und klassenbezogenen *Missinganalysen* unterzogen, die im Folgenden näher beschrieben werden. In dem Ausgangsdatsatz der *Schülerstichprobe* mit 1734 Fällen lagen 983 Fälle vor, die keinerlei Missings aufwiesen – weder bei den Items zur Beurteilung der Unterrichtsmerkmale noch bei denen zur Einschätzung der Urteilssicherheit. Wurden die zwei Itemtypen getrennt betrachtet, wiesen 1072 Fälle vollständige Datensätze bei den Items zur Einschätzung der Unterrichtsmerkmale und 1043 bei der Angabe der Urteilssicherheit auf. Von diesen Fällen wurden – bei getrennter Betrachtung der beiden Itemtypen – nur diejenigen Fälle eingeschlossen, bei denen mindestens 75% der geforderten Datenpunkte vorlagen, d. h. ein Fall durfte maximal 38 Missings aufweisen. Des Weiteren wurden Abbrecher gelöscht, die am Ende des Fragebogens durchgehend Missings enthielten. Auch Fälle, bei denen eine oder mehrere Fragebogenseiten nicht beantwortet wurden und dadurch die Itemeinschätzungen bei einem Unterrichtsmerkmal teilweise oder komplett fehlten, wurden entfernt. Die Anwendung dieser Kriterien führte zur Löschung von 47 Fällen. Fälle mit Extremwerten bzw. einer Häufung dieser wurden nicht entfernt, da diese auf ein tatsächlich ‚extremes‘ und somit informatives Antwortverhalten von Personen zurückgehen oder auf Instruktionsmissverständnisse hindeuten können. Das genaue Vorgehen bei der Datenbereinigung kann den Ausführungen auf der CD-Rom entnommen werden.

Hinsichtlich der klassenbezogenen Missinganalyse konnten aufgrund der Tatsache, dass bei jeder Lehrervariablen pro Klasse nur ein Datenpunkt vorliegt, einzelne Lehrkräfte nicht gelöscht werden. Eine Ausnahme stellten Lehrkräfte dar, die von mehreren Klassen beurteilt wurden und somit ein ‚Mehrfachfeedback‘ erhalten haben. Einem konservativen Vor-

gehen entsprechend, wird davon nur eine der erhobenen Klassen in die Analyse einbezogen, da ansonsten Konsistenzmaße artifiziell erhöht werden könnten. Die Entscheidungskriterien für die Auswahl der Klassen, die mehrfach erhoben wurden, sind die Schüleranzahl, die Anzahl gültiger Werte bei listenweiser Elimination und der prozentuale Missinganteil bei den Variablen zur Beurteilung der Unterrichtsmerkmale sowie zur Einschätzung der Urteilssicherheit. Dieses Vorgehen führte zur Entfernung von acht Schulklassen (Lehrer-ID 4, 9, 10, 11, 16, 30, 42 und 49, s. CD-Rom) und somit insgesamt 162 Fällen. Nach diesem Schritt stehen insgesamt 1525 Fälle mit 904 kompletten Fällen hinsichtlich der Itemangaben zur Erfassung der Unterrichtswahrnehmung als auch zur Einschätzung der Urteilssicherheit als effektive Stichprobe zur Verfügung. Die Stichprobenzusammensetzung in den einzelnen Phasen der Datenerhebung und -bereinigung kann Abbildung 9 entnommen werden.

Die dazugehörige Häufigkeitsverteilung der Missings bei Gegenüberstellung der beiden Itemtypen ist im Anhang C.2.1 zu finden. Legt man lediglich die Items zur Unterrichtswahrnehmung zugrunde, so sind insgesamt 986 vollständige Fälle vorhanden, die sich aus 491 Schülerangaben von insgesamt 773 Fällen in der ich-Version und 495 kompletten von insgesamt 752 Fällen in der wir-Version zusammensetzen. Somit weisen in der ich-Version 282, d. h. 36,5% der Probanden mindestens bei einem Item eine fehlende Angabe auf. Dieser Wert ist absolut gesehen relativ hoch, bedeutet aber, dass bei 260 Probanden mindestens 73 der 78 Items beantwortet wurden. In der wir-Version lassen sich 257 Fälle (34,2%) mit mindestens einer fehlenden Angabe identifizieren, wobei davon 229 Fälle weniger als 5 fehlende Werte aufweisen. Wird die Verteilung der Missings der Items zur Erfassung der Unterrichtswahrnehmung über die Klassen betrachtet, so weist eine Klasse in der Gesamtstichprobe im Durchschnitt nur 0.88 Missings (ich-Version: 0.91, wir-Version: 0.86) auf.

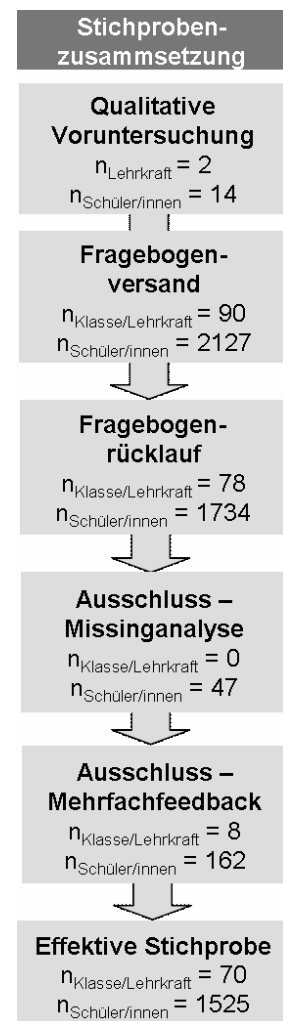


Abbildung 9: Stichprobenzusammensetzung innerhalb der verschiedenen Erhebungsphasen

### 7.3 Beschreibung der Stichproben

Die *Lehrkraftstichprobe* besteht aus 35 Lehrerinnen und 35 Lehrern, die im Durchschnitt seit 15,7 Jahren ihren Lehrberuf ausüben. Die Häufigkeitsverteilung der Dauer ihrer Lehrberufstätigkeit findet sich im Anhang C.2.2. Die Betrachtung der Schulartzuordnung zeigt, dass die Realschule mit 16 Lehrkräften und das Gymnasium sowie die Integrierte Gesamtschule mit jeweils 20 Lehrkräften am häufigsten und nahezu gleichverteilt vertreten sind. Ferner nehmen acht Hauptschullehrkräfte<sup>9</sup> und sechs Lehrkräfte der Regionalen Schulart an der Erhebung teil.

Tabelle 8: Verteilung der Lehrkräfte auf die Schulart und die Klassenstufe.

		Klassenstufe				
		7. Klasse	8. Klasse	9. Klasse	10. Klasse	n
Schulart	Hauptschule (HS)	3 (vs. 4)	1	1 (vs. 3)	3	8 (vs. 11)
	Realschule (RS)	6 (vs. 7)	5	1 (vs. 2)	4 (vs. 5)	16 (vs. 19)
	Gymnasium (Gym)	6	4	5	5	20
	Integrierte Gesamtschule (IGS)	3	4	2	11	20
	Regionale Schule (RegS)	2 (vs. 0)	0	3 (vs. 0)	1 (vs. 0)	6 (vs. 0)
	n	20	14	12	24	Σ 70

*Anmerkung:* Die Werte in Klammern stellen die geänderte Häufigkeitsverteilung dar, wenn die Zuteilung der Regionalen Schulklassen bezüglich des Haupt- und Realschulniveaus berücksichtigt und eine entsprechende Zuteilung zu der Haupt- und Realschule vorgenommen wird.

Hinsichtlich der Lehrkräfte an Regionalen Schulen wird aufgrund der geringen Anzahl auf die in dieser Schulform realisierte Leistungsdifferenzierung in Hauptschul- bzw. Realschulerniveau zurückgegriffen. Entsprechend dieser Leistungsdifferenzierung unterrichten diese Lehrkräfte in zwei reinen Hauptschul-, zwei reinen Realschulklassen sowie in zwei Klassen, welche die jeweiligen Lehrkräfte einmal als Hauptschul- und einmal als Realschulniveau einstufen<sup>10</sup>. Die Verteilung der befragten Klassenstufen auf die unterschiedlichen Schularten kann Tabelle 8 entnommen werden.

<sup>9</sup> An dieser Stelle soll darauf hingewiesen werden, dass die Hauptschule in Rheinland-Pfalz abgeschafft und ab dem Schuljahr 2009/10 der Hauptschulabschluss innerhalb der „Realschule plus“ angeboten wird, die unter ihrem Dach die beiden Schulformen Kooperative Realschule und Regionale Schule vereint (Ahnert, 2007).

<sup>10</sup> Der Rückgriff auf die Lernniveaueinschätzung der Lehrkraft erfolgt, weil entsprechend der Organisationsstruktur der beiden befragten Regionalen Schulen die Leistungsdifferenzierung für Deutsch erst ab der 8. Klasse und nicht wie für Englisch und Mathematik bereits ab der 7. Klasse erfolgt.

Hinsichtlich der erhobenen *Schulfächer* zeigt sich die in Tabelle 9 dargestellte Häufigkeitsverteilung, wobei Deutsch und Mathematik in allen Schularten und die Kernfächer Englisch, Biologie, Chemie und Physik mindestens in den Schularten Realschule, Gymnasium und der Integrierten Gesamtschule vertreten sind. Bei dem jeweilig beurteilten Unterrichtsfach handelt es sich nur bei zwei Lehrkräften um ein fachfremdes Unterrichtsfach.

Tabelle 9: Anzahl der eingeschätzten Unterrichtsfächer in Abhängigkeit von den verschiedenen Klassenstufen

		Klassenstufe				
		7. Klasse	8. Klasse	9. Klasse	10. Klasse	n
Unterrichtsfach	Deutsch	6	2	2	9	<b>19</b>
	Mathematik	6	1	6	3	<b>16</b>
	Englisch	2	4	2	1	<b>9</b>
	Französisch	1	3	0	2	<b>6</b>
	Biologie	1	0	0	3	<b>4</b>
	Chemie	0	2	1	2	<b>5</b>
	Physik	2	2	0	3	<b>7</b>
	Andere Fächer	2 (Latein, Gesellschaftslehre)	0	1 (Sozialkunde)	1 (Gesellschaftslehre)	<b>4</b>
	n	<b>20</b>	<b>14</b>	<b>12</b>	<b>24</b>	<b>Σ 70</b>

64% (n = 45) der 70 Lehrkräfte haben ein Lehramtsstudium in der Schulart absolviert, in der sie zum Zeitpunkt der Befragung unterrichten. 20 Lehrkräfte haben mit einer Hauptschul- (n = 3), Realschul- (n = 7) oder Gymnasialausbildung (n = 10) als Hintergrund die Integrierte Gesamtschule (n = 17) oder die Regionale Schule (n = 3) als Tätigkeitsfeld gewählt. Zwei Lehrkräfte unterrichten mit einer Ausbildung für die Grundschule bzw. das Gymnasium an einer Hauptschule und weitere drei Lehrkräfte sind Seiten- bzw. Quereinsteiger. Hinsichtlich der *Einschätzung der Unterrichtsstunde* geben 20% (n = 14) der Lehrkräfte an, dass die Unterrichtsstunde „etwas anders war als sonst“, die anderen (n = 56) stimmen der Aussage zu, dass die Unterrichtsstunde „genau so war wie sonst“. Der Großteil der Lehrkräfte unterrichtet die jeweils befragte Klasse im ersten oder zweiten *Jahr*. Eine genaue Verteilung der Häufigkeiten hinsichtlich des Unterrichtsjahres findet sich im Anhang C.2.3.

Die *Schülerstichprobe* umfasst 1525 Schüler auf 70 Klassen verteilt. Diese Klassen umfassen im Durchschnitt 22 Schüler bei einem Streubereich von [10; 30]. Die Verteilung der Klassen auf die Schulart, die Klassenstufen und das Unterrichtsfach entspricht der Darstel-

lung in der Tabelle 8 und Tabelle 9. 45% der Schüler ( $n = 690$ ) haben ihre letzte Note in dem beurteilten Fach angegeben. Der Notendurchschnitt über alle Fächer hinweg beträgt 2,98. Das Fach Biologie hat mit 2,54 ( $n = 24$ ) den besten und Chemie mit 3,60 ( $n = 45$ ) den schlechtesten Notendurchschnitt, allerdings müssen dabei verzerrende Effekte aufgrund der geringen zugrunde liegenden Stichprobenanzahl hinsichtlich der Notenbekanntgabe berücksichtigt werden. In Bezug auf die Klassenstufen zeigt sich bei einer ausreichend großen Notenanzahl der beste Notendurchschnitt in der 9. Klasse mit 2,77 ( $n = 104$ ) und der schlechteste in der 8. Klasse mit 3,20 ( $n = 127$ ). Die Schularten lassen sich hinsichtlich des Notendurchschnitts bei einer erneut stark variierenden Stichprobenanzahl in die folgende Rangreihe bringen: Hauptschule ( $\bar{x}_{\text{Note}} = 2,70$ ;  $n = 43$ ), Gymnasium ( $\bar{x}_{\text{Note}} = 2,82$ ;  $n = 244$ ), Integrierte Gesamtschule ( $\bar{x}_{\text{Note}} = 2,97$ ;  $n = 131$ ), Regionale Schule ( $\bar{x}_{\text{Note}} = 3,14$ ;  $n = 94$ ) und abschließend die Realschule ( $\bar{x}_{\text{Note}} = 3,20$ ;  $n = 178$ ). Weitere demographische Daten zur Schülerstichprobe sind nicht verfügbar, da in Absprache mit dem Ministerium für Bildung, Wissenschaft, Jugend und Kultur in Rheinland-Pfalz sowie dem Datenschutzbeauftragten Rheinland-Pfalz sozialstatistische Schülerdaten bei der Haupterhebung nicht erhoben wurden, um eine anonyme Befragung zu gewährleisten.

#### 7.4 Ergebnisrückmeldung an die Lehrkräfte

Nach der Datenaufbereitung erfolgte zeitnah die Ergebnisrückmeldung an die Lehrkräfte. Im Abschnitt 4.2 wurden bereits Überlegungen zur Gestaltung von Ergebnisrückmeldungen von Evaluationsergebnissen skizziert. Diese fließen als Forderung nach einer verständlichen, übersichtlichen und motivierenden Gestaltung der im Projekt verschickten Ergebnisrückmeldungen ein.

Der Ergebnisbericht wird im pdf-Format per E-Mail an eine vorher von der Lehrkraft genannte Adresse versendet, um sicherzustellen, dass ausschließlich die jeweilige Lehrkraft auf den Ergebnisbericht zugreifen kann. Ein anonymisiertes Exemplar der insgesamt 78 verschickten Ergebnisberichte findet sich im Anhang C.3. Die graphische Aufbereitung (s. Anhang C.3.2) wurde von Wolfgang Wagner übernommen.

Die Schülerangaben werden aggregiert, auf Itemebene und vorzugsweise graphisch zurückgemeldet. Des Weiteren werden lediglich grundlegende deskriptive Kennwerte angegeben: Der Klassenmittelwert sowie die -streuung (numerisch), die Häufigkeitsverteilung der Schülereinschätzungen (numerisch und graphisch), der Prozentsatz zustimmender („stimme

eher zu' und ‚stimme zu') Schülereinschätzungen (numerisch) und die Stichprobengröße pro Item (numerisch). Hinsichtlich des Vergleichs mit der Selbsteinschätzung der Lehrkraft wird das Lehrkraft- und Schülerurteil in einem Verlaufsprofil über alle Items hinweg graphisch dargestellt und der Differenzwert zwischen den beiden Angaben numerisch angegeben. Die aufgeführten Kennwerte werden in einem von der Autorin verfassten Anschreiben erläutert und mit Hinweisen zur Interpretation ergänzt. Es wird auch erwähnt, dass die Fragebogenerhebung in diesem ersten Schritt auf die Überprüfung der Eignung der Items abzielt und der Fragebogen deswegen z. T. inhaltlich ähnliche Items enthält, wie z. B. Item 4 „In der Stunde ging es nur um Dinge, die mit dem Lernstoff zu tun haben“ und Item 11 „Es wurde keine Zeit für Dinge verwendet, die mit dem Lernstoff nichts zu tun haben“). Ferner wird darauf hingewiesen, dass bei einer Fragebogenerhebung Messfehler grundsätzlich nie ausgeschlossen werden können. Sowohl instrumentelle Schwächen aber auch äußere (z. B. Störgeräusche) und innere (z. B. Stimmungswechsel) situative Bedingungen können zu Über- oder Unterschätzungen führen. Erhebungsergebnisse sind daher lediglich als eine Anregung zur Reflexion und Verbesserung des Unterrichts zu sehen und die Ergebnisse müssen sorgfältig geprüft werden, bevor Konsequenzen daraus abgeleitet werden dürfen. Obgleich die Gestaltung dieses Prüfprozesses nur von den Lehrkräften selbst durchgeführt werden kann, da sie die Experten für ihre Klasse und das konkrete Unterrichtsgeschehen in der beurteilten Unterrichtsstunde sind, werden diesbezüglich unterstützende Anregungen gegeben, z. B. in der Form, dass unerwartet große Diskrepanzen zwischen den Urteilsperspektiven zu prüfen sind.

## 8 Empirische Vorarbeiten

Im Abschnitt 8.1 soll ein kurzer Überblick über das Vorgehen bei der Auswertung der Fragebogendaten zur Orientierung gegeben werden. Genauere Ausführungen zu den jeweiligen Auswertungsmethoden und relevante Hintergründe werden aufgrund der besseren Übersichtlichkeit dem jeweiligen Ergebnisabschnitt vorangestellt. Inwiefern die Daten die jeweiligen Analysevoraussetzungen erfüllen, ist Gegenstand des Abschnitts 8.2.

### 8.1 Auswertungsvorgehen

Die Items zur Unterrichtswahrnehmung werden anhand verschiedener statistischer Verfahren im Hinblick auf ihre Eignung zur Erfassung der Unterrichtsmerkmale analysiert. Dazu erfolgt im ersten Schritt eine deskriptivstatistische Analyse der Items zur Unterrichtswahrnehmung mit Blick auf die Gesamtdaten und jeweils getrennt für die ich- und wir-Version des Fragebogens. Im nächsten Schritt werden gemäß einem klassisch test-theoretischen Vorgehen Item- und Reliabilitätsanalysen durchgeführt. Da die klassischen Kennwerte aufgrund ihrer Abhängigkeit von einer Reihe von Faktoren, wie z. B. dem Itemmittelwert, der Itemverteilung oder der Standardabweichung zur Fragebogengüte nur eine begrenzte Aussagekraft haben, werden in einem nächsten Schritt faktorenanalytische Verfahren zur Beurteilung der Dimensionalität durchgeführt. Die Überprüfung der faktoriellen Validität erfolgt mittels eines explorativen und konfirmatorischen Zugangs. Die exploratorische Faktorenanalyse zielt auf eine mögliche Datenreduktion und die Ermittlung vorhandener Datenstrukturen ab. Bei der konfirmatorischen Analyse werden die Unterrichtsmerkmale als latente Konstrukte postuliert und die Items als Indikatorvariablen auf diese zurückgeführt. Im Modelltest wird die Nullhypothese geprüft, inwiefern die empirische Varianz-Kovarianzmatrix der modelltheoretischen Varianz-Kovarianzmatrix des postulierten spezifizierten Modells entspricht. Die Entscheidung, ob ein Item in der jeweiligen Skala verbleibt oder nicht, sollte allerdings nicht allein von statistischen Kennwerten zur Beurteilung der Fragebogengüte abhängig sein. Homogenitätskennwerte zielen z. B. ausschließlich auf die Homogenität der Skala ab, wobei innerhalb einer Skala nicht mit jedem Item der gleiche Inhalt gemessen werden sollte. Deswegen fließen auch Aspekte der Inhalts- und Augenscheinvalidität sowie praktische Erfordernisse als Kriterium hinsichtlich der endgültigen Itemselektion ein. Die Eignung der ich- vs. wir-Formulierung erfolgt dabei anhand einer Zusammenstellung der aufgrund dieser Analysen gewonnenen Ergebnisse in

Abwägung mit den im Abschnitt 3.4.1 aufgeführten Überlegungen zu den stattfindenden Urteilsprozessen sowie ihren Konsequenzen. Die Items zur Urteilssicherheit werden separat hinsichtlich ihres Informationsgehalts analysiert.

Alle Analysen werden mit Hilfe der Softwarepakete SPSS 14.0, SYSTAT 10.2, AMOS 6.0 sowie EQS 6.1 durchgeführt. Die Antwortkategorien werden folgendermaßen kodiert: „stimme nicht zu“ = 1, „stimme eher nicht zu“ = 2, „stimme eher zu“ = 3 und „stimme zu“ = 4. Im folgenden Abschnitt wird zunächst geprüft, inwiefern die Voraussetzungen zur Durchführung der intendierten Analysen vorliegen.

## 8.2 Prüfung der Analysevoraussetzungen

Im Folgenden werden relevante Einflussfaktoren der Datenqualität bezüglich der durchzuführenden Analysen betrachtet und hinsichtlich des vorliegenden Datensatzes beurteilt.

*Skalenniveau.* Den anhand einer vierstufigen Antwortskala erfassten Unterrichtswahrnehmungen und Einschätzungen der Urteilssicherheit wird Intervallskalenniveau unterstellt. Diese Annahme wird dadurch gestützt, dass 79% der Befragten in der qualitativen Voruntersuchung die Antwortkategorien als symmetrisch wahrnehmen. Auch zeigen Simulationsstudien im Kontext von Strukturgleichungsmodellen, dass das Ausmaß von potentiellen Parameter- und Standardfehlerverzerrungen unter diesen Bedingungen dann gering ist, wenn die Variablen bezüglich der univariaten Schiefe- und Exzesswerte im Bereich zwischen -1.0 und +1.0 liegen (Muthén & Kaplan, 1992, zitiert nach Wagner, 2008). Im Schülerdatensatz zeigt sich, dass nur 18 Items (ich-Version: 21 Items, wir-Version: 20 Items) außerhalb des von Muthén und Kaplan (1992) geforderten Bereichs liegen. Und davon weisen lediglich die zwei Items SLK22 und SKV46 einen Schiefewert größer 2.0 auf. Hinsichtlich der Exzesswerte liegen in der Gesamtstichprobe und ich-Version 33 Items (wir-Version: 36 Items) außerhalb dieses Bereichs, wovon wiederum nur fünf Items (ich-Version: vier Items) einen Exzess-Wert größer 2.0 aufweisen. Hinsichtlich der hohen Itemanzahl von 78 und geringen deutlichen Ausreißern kann dieses Kriterium daher als erfüllt angesehen werden.

*Stichprobengröße.* Mummendey, Mendoza, Stafford und Stauffer (2000) geben an, dass die Stichprobengröße für zuverlässige Reliabilitätsschätzungen mindestens 100 betragen sollte. Mit 1525 Schülern insgesamt (ich-Version: 773, wir-Version: 752), davon 904 mit vollständigen Angaben (ich- und wir-Version: 452) wird dieser Wert auf der Individualebene in



der vorliegenden Stichprobe weit überschritten. Obwohl die Analyse der ich- und wir-Version quasi zur Teilung der Stichprobe auf Klassenebene führt, reicht die vorliegende Fallzahl für die Analysen aus. Die Berechnung von Klassenmittelwerten auf der Datengrundlage der Klassenhälfte stellt überdies ein gängiges Vorgehen in Studien, wie z. B. PISA dar.

Mit 70 Stichprobenelementen auf der Klassenebene ist dieses Kriterium nicht erfüllt. Da der Analysefokus jedoch auf der Individualdatenebene liegt und Analysen, in die die Klassenebene einbezogen werden nur in geringfügigem Umfang durchgeführt werden, ist die Klassengröße akzeptabel. Aufgrund dessen, dass zwei verschiedene Versionen mit Hilfe der Stichprobe überprüft werden, wird von einer Teilung der Stichprobe zur Kreuzvalidierung abgesehen, da dies die den Stichprobenumfang erheblich verringern würde.

Die Stichprobengröße bietet somit exzellente Voraussetzungen für die Durchführung einer Faktorenanalyse. Zwar zeigten MacCallum, Widaman, Zhang und Hong (1999), dass bei Itemkommunalitäten ( $h^2$ ) von mindestens .60 bereits 60 Stichprobenelemente ausreichen, aber generell erhöht sich mit der Zunahme der Stichprobengröße die Stabilität der Faktorenlösung.

*Umgang mit Missings.* Der in Abschnitt 7.2 angegebene relativ hoch erscheinende Anteil fehlender Angaben in der ich- und wir-Version ist hinsichtlich der Datenanalysen nicht problematisch, da sich die Missings relativ gleichmäßig über die Klassen und Items verteilen. Dennoch wird ein Imputationsverfahren zur Absicherung herangezogen. Der Vergleich zwischen EM-imputierten Daten und den Daten ohne Imputation erbringt, dass sich Mittelwerts- und Standardabweichungsunterschiede insgesamt und in der ich- bzw. wir-Version bis auf geringfügige Ausnahmen erst in der dritten bzw. vierten Nachkommastelle zeigen und somit vernachlässigbar sind. Die Rangfolge der Reliabilitätsanalysekennwerte bleibt ebenfalls gleich. Um die Vergleichbarkeit mit den deskriptiven Daten zu gewährleisten wird daher das konservativere Vorgehen ohne die Missingimputation bei den Reliabilitätsanalysen gewählt. Um bei der exploratorischen und konfirmatorischen Faktorenanalyse Schätzprobleme zu vermeiden, wird hier allerdings auf die Full Information-Maximum-Likelihoodschätzung (FIML) zurückgegriffen, die das Likelihood-Modell unter Berücksichtigung der vorhandenen Missings maximiert und sich asymptotisch dem EM-Imputationsverfahren annähert (Arbuckle, 1996). Die zur Analyse verwendeten Korrelationsmatrizen werden anhand der Software SAS mit Hilfe der MI-Prozedur von Wolfgang

Wagner erstellt. Diese Korrelationsmatrizen stellen eine gute Annäherung dar, optimal wären Matrizen auf der Grundlage von polychorischen Korrelationen, deren Berechnung ist aber bei dem Vorhandensein von Missings mit einem deutlich höheren Aufwand verbunden.

*Normalverteilung.* Eine Normalverteilung der Items kann aufgrund dessen, dass es sich um Variablen mit nur vier Kategorien handelt, bereits theoretisch zurückgewiesen werden. Die Überprüfung gemäß Brosius (2002) anhand von Normalverteilungsdiagrammen, welche die empirisch beobachteten Werte gegen die unter einer Normalverteilungsannahme theoretisch erwarteten Werte plottet, bestärkt diese Annahme – die Items weisen durchgängig Abweichungen von der Normalverteilung auf. Auch der Kolmogorov-Smirnov-Test, der insbesondere für große Stichproben gut geeignet ist, weist die Annahme der Normalverteilung in der Grundgesamtheit mit einem  $p \leq .000$  durchgängig zurück. Inferenzstatistisch zeigt sich, dass alle Items entweder hinsichtlich der Schiefe oder der Kurtosis von null abweichen. Allerdings liegen bis auf die beiden bereits erwähnten Items SLK22 und SKV46, alle übrigen Items innerhalb der von West, Finch und Curran (1995) angegebenen Grenze einer Kurtosis  $< 7$  und einer Schiefe  $< 2$ . Auf eine Logarithmierung zur Normalisierung der Itemverteilung soll aus Gründen der fehlenden Vergleichbarkeit der transformierten Items mit den Originalitems verzichtet werden, da diese Vergleichbarkeit im Rahmen der Konstruktion eines Fragebogens essentiell ist. Wird auf Bühner (2006) rekuriert, ist das nicht nötig, da bei der Durchführung einer Hauptkomponentenanalyse, die im vorliegenden Fall bei der exploratorischen Faktorenanalyse gewählt wird, normalverteilte und intervallskalierte Items zwar optimal, aber aufgrund der relativ hohen Robustheit der Schätzverfahren nicht zwingend notwendig sind.

Zum Prüfen des Vorliegens einer multivariaten Normalverteilung hinsichtlich der Durchführung von konfirmatorischen Faktorenanalysen wird der Mardia-Test (Mardia, 1970) herangezogen. Dieser weist einen Wert von 837,38 bei einem „critical-ratio“-Wert von 146,36, welcher als z-Wert interpretiert werden kann. Bereits ein z-Wert von 2,58 entspräche bei zweiseitiger Signifikanztestung einem p-Wert von  $< .001$ . Somit ist die Annahme der multivariaten Normalverteilung im vorliegenden Fall deutlich verletzt. Hinsichtlich möglicher multivariater Ausreißerwerte, die aufgrund der mit AMOS ermittelten Mahalanobis Differenzwerte bestimmt werden können und die Distanz eines Falles vom Schwerpunkt (Zentroid) aller beobachteten Datenpunkte angeben, liegen keine ungewöhnlichen

Werte vor. Allerdings ist diese Prüfmethode stichprobenabhängig und von daher hinsichtlich seiner Aussagekraft eingeschränkt.

Aufgrund der Verletzung der multivariaten Normalverteilungsannahme müssten die Maximum Likelihood-Methode (ML) und Generalized Least Squares-Schätzmethode (GLS) bei einem strengen Vorgehen ausgeschlossen werden (Bollen, 1989). Schätzverfahren, die keine multivariate Normalverteilung voraussetzen, sind der Scale Free Least-Squares-Schätzer (SLS) und der Unweighted Least Squares-Schätzer (ULS) (Backhaus, Erichson, Plinke & Weiber, 2006). Der SLS-Schätzer erlaubt allerdings keine inferenzstatistischen Schlüsse (Backhaus et al., 2006) und die ULS-Methode ist nicht skalenfrei, skaleninvariant und nimmt keine effiziente Schätzung der Diskrepanzfunktion vor (Bühner, 2006). Die diese Eigenschaften aufweisende Asymptotic Distribution Free-Methode (ADF) ist allerdings bei den zu überprüfenden postulierten Modellen in der Mehrheit nicht anwendbar, da der Stichprobenumfang zu gering ist. Die ADF-Schätzmethode fordert bei  $p \geq 12$  manifesten Variablen einen Stichprobenumfang von  $1,5p(p+1)$  Fällen, d. h. im vorliegenden Fall des Gesamtmodells mit 78 manifesten Variablen 9243 Stichprobenelemente. Bühner (2006) empfiehlt in diesem Fall auf die Bollen-Stine-Bootstrap-Korrektur zurückzugreifen. Deshalb wird die ML-Schätzung bei den konfirmatorischen Faktorenanalysen angewendet, aber in Ergänzung mit dem Bootstrapping-Verfahren, welches Über- oder Unterschätzungen aufgrund der verletzten multivariaten Normalverteilungsannahme identifiziert (Byrne, 2001)

*Linearzusammenhänge.* Eine wichtige Voraussetzung für die Durchführung einer Faktorenanalyse sind nach Bühner (2006) ausreichend hohe lineare Zusammenhänge zwischen den Items. Die Analyse der Korrelationsmatrix der Items zur Erfassung der Unterrichtswahrnehmung zeigt die in Tabelle 10 dargestellte Verteilung der insgesamt  $(78 \text{ Items} (78 \text{ Items} - 1)) / 2 = 3003$  Produktmomentkorrelationen, die die Grundlage für die Faktorenanalyse darstellen. Gemäß der Klassifikation von Cohen (1992) weist die Mehrzahl der Items eine positive Korrelation kleiner Effektstärke auf und immerhin noch 15% der Korrelationen ( $n_{\text{Korrelation}} = 458$ ) lassen sich als mittlere Effektstärke einstufen. Der Kaiser-Meyer-Olkin-Koeffizient (KMO) zeigt, dass mit einem Wert von .957 substantielle Korrelationen vorliegen, die die Durchführung einer Faktorenanalyse rechtfertigen. Der KMO-Koeffizient setzt den gemeinsamen Varianzanteil aller Items zu der Summe aus diesem und den quadrierten Partialkoeffizienten der Items ins Verhältnis und wird groß,

wenn die Items nur über geringe spezifische Varianzanteile verfügen (Bühner, 2006). Auch der Bartlett-Test, der die Nullhypothese einer Nullkorrelation aller Items prüft, ist bei einem  $\chi^2$ -Wert von 40046,575 bei 3003 Freiheitsgraden hochsignifikant. Dabei gilt es allerdings zu bedenken, dass diese Signifikanz maßgeblich auf die große Stichprobenanzahl und die damit einhergehende hohe Teststärke zurückzuführen ist. Allerdings sind auch die Anti-Image-Werte unter- und oberhalb der Diagonalen, d. h. die Partialkorrelationen nach Auspartialisierung aller anderen Items nahezu 0. Somit ist der von den anderen Items unabhängige Varianzanteil gering, und es kann nahezu die gesamte gemeinsame Varianz zwischen zwei Items durch die übrigen Items aufgeklärt werden.

Einen Anhaltspunkt für die Eignung der einzelnen Items für eine Faktorenanalyse bietet der MSA-Koeffizient (Measure of Sample Adequacy), der die Partialkorrelation eines Items mit den übrigen Items ins Verhältnis setzt und möglichst hoch sein sollte. Bis auf die Items SKF2 und SLK19 weisen alle Items gemäß der Klassifikation von Bühner (2006) mit MSA-Koeffizienten  $> .90$  eine sehr gute Eignung für die Faktorenanalyse auf. Die beiden erwähnten Items weisen eine mittlere bzw. gute Eignung auf ( $MSA_{SKF2}$ -Koeffizient = .79,  $MSA_{SLK19}$ -Koeffizient = .80).

Tabelle 10: Verteilung der Produktmomentkorrelationen innerhalb der gesamten Schülerstichprobe

Korrelationsintervall	$-.13 \leq r < 0$	$0 \leq r < .10$	$.10 \leq r < .30$	$.30 \leq r < .50$	$.50 \leq r < .69$	$\Sigma$
$n_{\text{Korrelation}}$	13	276	2246	458	10	3003

Bezüglich der Durchführung von Faktorenanalysen sind die Voraussetzungen somit als gut einzustufen. Minderungen der Korrelationen können u. a. durch Abweichungen der Items von der Normalverteilung verursacht werden. Die Items besitzen unterschiedliche Schiefen, was dazu führt, dass die Items nicht maximal miteinander korrelieren können. Eine weitere Möglichkeit, die zur Minderung der Korrelation führen könnte, wären Ausreißerwerte. Diese haben beim vorliegenden Datensatz jedoch keinen substantiellen Einfluss.

*Multikollinearität.* Bühner (2006) rät ab einer Korrelation von .85 zum Ausschluss der Items aufgrund einer vorliegenden Multikollinearität, die zu Schätzproblemen bei der konfirmatorischen Faktorenanalyse führen könnte. Die Spannweite der Korrelationen reicht von  $r_{SKF5, SLK19} = -.13$  bis  $r_{SKF7, SKF13} = .69$ , somit deutet auch der höchste Korrelationskoeffizient nicht auf eine ernsthafte Multikollinearität hin.

## 9 Analyse der Items zur Unterrichtswahrnehmung

In diesem Kapitel werden die Items zur Erfassung der Unterrichtswahrnehmung analysiert. Nach einer deskriptivstatistischen Betrachtung (Abschnitt 11.1) folgen die ausführlichen Item- und Reliabilitätsanalysen (Abschnitt 11.2) sowie die Überprüfung der faktoriellen Validität (Abschnitt 11.3) und die Betrachtung weiterer Validitätskriterien (Abschnitt 11.4). Die sich aufgrund der Analyseergebnisse ergebenden Schlussfolgerungen werden jeweils am Ende des Abschnitts aufgeführt.

### 9.1 Deskriptivstatistische Befunde

Im Folgenden werden die Unterrichtswahrnehmungsitems nach einem kurzen Überblick zum geplanten Auswertungsvorgehen (Abschnitt 9.1.1) hinsichtlich ihrer Missinganzahl (Abschnitt 9.1.2), ihrer Verteilungsform (Abschnitt 9.1.3), ihrer Lokation (Abschnitt 9.1.4) und Dispersion (Abschnitt 9.1.5) analysiert. Eine Übersicht zu den deskriptiven Kennwerten findet sich im Anhang D.1.1. Der Abschnitt 9.1.6 führt wesentliche Befunde überblicksartig zusammen.

#### 9.1.1 Auswertungsvorgehen

Zunächst werden die *Missingverteilungen* pro Item betrachtet, wobei Analysen auf Klassenebene hinzugezogen werden, da es durchaus vorkommen kann, dass es viele fehlende Werte bei einem Item gibt, diese sich aber im Wesentlichen auf eine kleine Klassenanzahl konzentrieren. Anschließend wird die *Verteilungsform* der Items hinsichtlich der Schiefe als drittes zentrales Moment und dem Exzess als viertes zentrales Moment fokussiert. Im nächsten Schritt beschreiben die *Lokationsmaße* als Maße der zentralen Tendenz die Lage der Verteilung. Ausgehend von der Intervallskalenniveauannahme wird dabei der Mittelwert betrachtet und der Median vergleichend hinzugezogen, da dieser bei einer konservativen Ordinalskalenniveauannahme streng genommen angesetzt werden müsste (Bühner, 2006). Anhand der *Dispersionsmaße* wird die Variabilität der beobachteten Messwerte bzw. deren Tendenz nicht den zentralen Tendenzwerten zu entsprechen, betrachtet. Analysiert werden dabei vorrangig die Kennwerte Minimum, Maximum, relative Kategorienhäufigkeit sowie die Standardabweichung. Es erfolgt dabei eine separate Darstellung der Befunde für die Gesamtstichprobe und bei entsprechenden Abweichungen separat für die ich- und wir-Version des Schülerfragebogens.

### 9.1.2 Itemmissinganalyse

Im Modus des paarweisen Fallausschlusses sind in der Gesamtstichprobe hinsichtlich der Items zur Unterrichtswahrnehmung im Durchschnitt 17 Missings pro Item (ich-Version: 9 Missings; wir-Version: 8 Missings) vorhanden. Somit liegen zwischen 1434 valide Schülerangaben (ich-Version: 728; wir-Version: 706) bei dem Item SKF10 und 1521 beim Item SKF17 (ich-Version: 772, Item SA68; wir-Version: 751, Items SKF17 und SLK18) vor. Dies entspricht einem sehr geringen Anteil fehlender Werte von 0.3% (ich- und wir-Version: 0.1%) bis maximal 6.0% (ich-Version: 5.8%; wir-Version: 6.1%). Dabei haben die Items SKF17, SLK18 und SA68 mit nur vier Missings insgesamt sowie in den beiden Fragebogenversionen ( $\text{Missings}_{\text{ich-Version}}(\text{SA68}) = 1$ ;  $\text{Missings}_{\text{wir-Version}}(\text{SKF17, SLK18}) = 1$ ) die geringste Missinganzahl. Die beiden Items mit den meisten Missings sind SKF10 („Bestimmte Aufgaben sind von dazu eingeteilten Schülern/innen (z. B. Klassendienste) erledigt worden“,  $\text{Missings}_{\text{insgesamt}} = 91$ ,  $\text{Missings}_{\text{ich-Version}} = 45$ ,  $\text{Missings}_{\text{wir-Version}} = 46$ ) und Item SM36 („Die Arbeitsblätter haben mir/uns Spaß gemacht“,  $\text{Missings}_{\text{insgesamt}} = 75$ ,  $\text{Missings}_{\text{ich-Version}} = 37$ ,  $\text{Missings}_{\text{wir-Version}} = 38$ ).

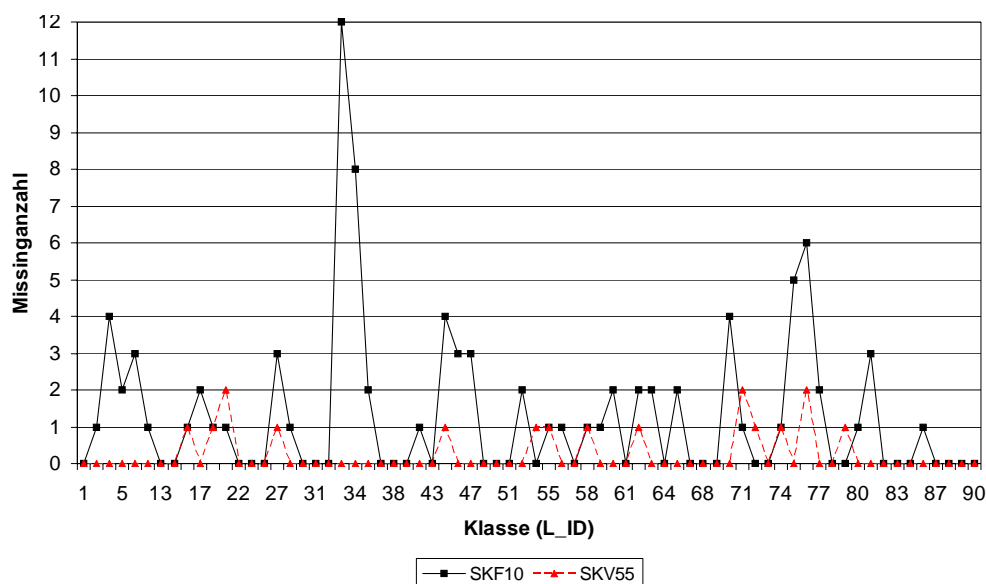


Abbildung 10: Relation der Missinganzahl des Items SKF10 zu der von Item SKV55 mit einer durchschnittlichen Anzahl von 17 Missings über alle erfassten Klassen hinweg

Setzt man die Missinganzahl der beiden Items SKF10 und SM36 hinsichtlich möglicher Schlussfolgerungen für die Itemauswahl in Relation zu der Gesamtstichprobe von 1525 Schülern/innen, ist diese vernachlässigbar. Betrachtet man diese allerdings in Relation zur durchschnittlichen Missinganzahl aller Items ( $n_{\text{Missing}} = 17$ ), so weisen die beiden Items durchschnittlich ein Missing mehr pro Klasse auf. In Abbildung 10 ist dieser Sachverhalt

für das Item SKF10 verdeutlicht. Dargestellt ist der Vergleich der Missinganzahl des Items SKF10 mit der des Items SKV55, welches entsprechend dem Durchschnitt 17 Missings aufweist. Für das Item SM36 ergibt sich ein ähnliches Muster. Die erhöhte Missinganzahl der Items SKF10 und SM36 sind mit hoher Wahrscheinlichkeit auf ihren konditionalen Charakter zurückzuführen. Trifft der zu beurteilende Sachverhalt nicht zu, ist eine Einschätzung nicht möglich. Dass dennoch nicht alle Schüler der Klasse ein Missing aufweisen, lässt sich u. a. durch den Rückgriff auf die Einschätzung dieses Sachverhalts im Allgemeinen bei dieser Lehrkraft bzw. auf die vergangenen Beispiele erklären. Aufgrund dieser Befunde liegt der Ausschluss der beiden Items SKF10 und SM36 nahe.

### 9.1.3 Verteilungsanalyse

Von der Prüfung auf Normalverteilung (s. Abschnitt 8.2) ist bekannt, dass keines der Items mit statistischer Signifikanz normalverteilt ist. Die Betrachtung der *Schiefewerte* zeigt, dass mit 63 Items (ich-Version: 63 Items, wir-Version: 65 Items) der Großteil der Items sowohl in der ich- als auch in der wir-Version numerisch eine negative Schiefe aufweist, welche Linksschiefe impliziert. Die restlichen 15 Items weisen numerisch eine positive Schiefe und somit eine Rechtsschiefe auf. Die Spannweiten der Schiefewerte in Abhängigkeit von der jeweiligen Version können Tabelle 11 entnommen werden.

Tabelle 11: *Spannweiten negativer und positiver Schiefewerte für die Fragebogenversionen sowie insgesamt*

Fragebogen- version	Spannweite	
	negativer Schiefewerte	positiver Schiefewerte
insgesamt	-0.04 (SKF3) bis -2.34 (SLK22)	0.02 (SA75) bis 1.01 (SA77)
ich-Version	-0.03 (SKF3) bis -2.27 (SLK22)	0.01 (SM43) bis 1.29 (SA72)
wir-Version	-0.01 (SA75) bis -2.40 (SLK22)	0.04 (SM38) bis 0.98 (SA77)

Das Item SLK22 „Der Lehrer hat die Klasse zu Beginn der Stunde freundlich begrüßt“ weist durchgängig die stärkste Linksschiefe auf. Die Schüler haben diesem Item somit in hohem Maße zugestimmt. Wird die *Kurtosis* bzw. der *Exzess* herangezogen, zeigt sich, dass der überwiegende Teil der Items eine Kurtosis kleiner 0 aufweist und somit eine gegenüber der Normalverteilung breitgipfligere bzw. flachere Verteilung hat. Nur 18 Items (ich-Version: 16 Items; wir-Version: 22 Items) weisen eine Kurtosis größer 0 auf, was auf eine eher schmalgipflige Verteilung mit einem steileren Gipfel als bei der Normalverteilung hindeutet. Die entsprechenden Spannweiten der Kurtosis finden sich in Tabelle 12.

Tabelle 12: *Spannweiten negativer und positiver Kurtosiswerte für die Fragebogenversionen sowie insgesamt*

Fragebogen- version	Spannweite	
	negativer Kurtosiswerte	positiver Kurtosiswerte
insgesamt	-1.56 (SA75) bis -0.04 (SKF14)	0.05 (SA62) bis 5.33 (SLK22)
ich-Version	-1.58 (SA75) bis -0.05 (SKF14)	0.24 (SKF1) bis 5.16 (SLK22)
wir-Version	-1.60 (SKF12) bis -0.03 (SKF14)	0.01 (SLK25) bis 5.16 (SLK22)

Die im Vergleich zur Normalverteilung überwiegend flachere Itemverteilung lässt sich u. a. auf den dichotomen Grundcharakter vieler Items zurückführen. So fordert die Beurteilung des Items SLK22, inwiefern die Lehrkraft die Klasse freundlich begrüßt hat, eher eine Ja/Nein-Entscheidung als eine abstufende Zustimmung. Weisen die Schülerurteile diesbezüglich eine geringe Übereinstimmung auf, trägt dies zur Verflachung der Verteilung bei. Besteht hingegen eine hohe Übereinstimmung der Schülerurteile, liegt eine gegenüber der Normalverteilung steilere Verteilung vor – dies ist im Beispiel des Items SLK22 der Fall.

Keines der Items hat eine Schiefe und eine Kurtosis von 0, was bei einer symmetrischen Verteilung der Fall wäre. Da die Angabe von Schiefe und Kurtosis nicht ausreicht, um irreguläre Verteilungsformen von verzerrten Normalverteilungen zu unterscheiden, wird die Verteilungsform zusätzlich über Histogramme<sup>11</sup> ermittelt. In Abbildung 11 ist jeweils ein Beispiel pro ermittelter Verteilungsform angegeben.

Die visuelle Analyse kommt bei 48 Items ebenfalls zur Einstufung als linksschiefe Items. Und sieben Aktivierungitems (SA61, SA69, SA70, SA71, SA72, SA73 und SA77) können auch anhand der Histogramme als rechtsschief eingestuft werden. Bei den anderen 27 Items führt die Analyse zu einer anderen Einschätzung der Verteilungsform. So weisen die Items SKF5, SKF7, SKF13, SKF17, SM37 und SM44 tendenziell jeweils verzerrte Normalverteilungen auf. Die Items SKF3, SKF10, SM32, SM35, SM36, SM38, SM39, SM40, SM43, SKV45, SKV52, SKV57, SA66, SA67 und SA74 nähern sich einer Gleichverteilung an. SKF12 und SA75 sind zweigipflig in U-Form verteilt, bei denen die beiden extremen Antwortkategorien gleichermaßen überproportional vertreten sind.

Die *ich-* und *wir-Version* unterscheidet sich hinsichtlich der Schiefe und Kurtosis bei wenigen Items im Vorzeichen (s. Anhang D.1.2). Hinsichtlich der Verteilungsform zeigen

<sup>11</sup> Die Autorin ist sich der Abhängigkeit der Histogramm-Interpretation von der jeweiligen Intervall- und Skalenbildung und der damit einhergehenden potentiellen Fehlerbehaftetheit der visuellen Verteilungseinschätzung bewusst und prüft deshalb bei nicht eindeutigen Items zusätzlich verschiedene Skaleneinstellungen.



sich Unterschiede bei zwei Items: Item SA74 weist in der ich-bezogenen Formulierung eine rechtsschiefe Verteilung auf, nähert sich in der wir-bezogenen Formulierung (und in der Gesamtstichprobe) aber einer Gleichverteilung an. Und Item SA78 ist in der ich-Version (und in der Gesamtstichprobe) deutlich linksschief, in der wir-Version allerdings gleichverteilt.

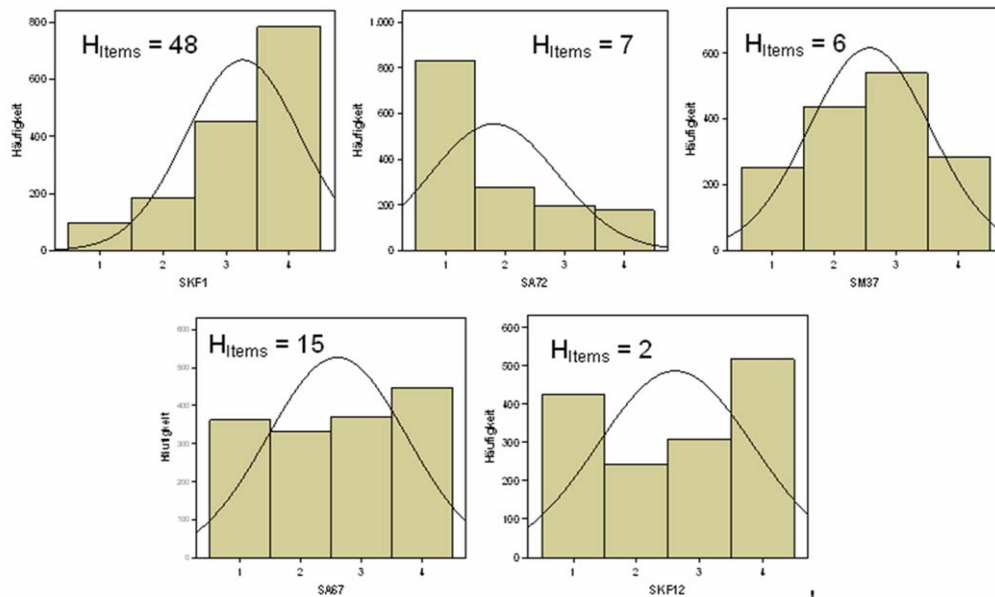


Abbildung 11: Histogramme zum linksschiefen Item SKF1, rechtsschiefen Item SA72, zum eine verzerrte Normalverteilungsform aufweisenden Item SM37, zum gleichverteilten Item SA67 und zum zweigipflig verteilten Item SKF12 sowie Angabe der Häufigkeiten der jeweiligen Verteilungsformen in der Gesamtstichprobe ( $H_{Items}$ )

Hinsichtlich der Schlussfolgerungen für die Itemauswahl lässt sich festhalten, dass die zweigipfligen Verteilungen der Items SKF12 und SA75 dann einen Itemausschluss erfordern, wenn man als Ursache ein unterschiedliches Frageverständnis über die Klassen hinweg, aber auch innerhalb der Klassen annimmt. Die Betrachtung der Verteilungsform innerhalb der einzelnen Klassen zeigt bei beiden Items eine große Varianz. Die beiden Items weisen innerhalb der Klassen sowohl linksschiefe ( $n_{Klassen}$  (SKF12) = 20,  $n_{Klassen}$  (SA75) = 21), rechtsschiefe ( $n_{Klassen}$  (SKF12) = 22,  $n_{Klassen}$  (SA75) = 25) und verzerrte Normalverteilungen ( $n_{Klassen}$  (SKF12) = 7,  $n_{Klassen}$  (SA75) = 2), als auch zweigipflige ( $n_{Klassen}$  (SKF12) = 13,  $n_{Klassen}$  (SA75) = 16), gleichverteilte ( $n_{Klassen}$  (SKF12) = 3,  $n_{Klassen}$  (SA75) = 5) und Verteilungen auf, bei denen nur eine oder zwei Antwortkategorien besetzt sind ( $n_{Klassen}$  (SKF12) = 5,  $n_{Klassen}$  (SA75) = 1). Setzt man die Häufigkeit irregulärer Verteilungsformen (zweigipflige Verteilungen) ins Verhältnis zur Gesamtanzahl auftretender Verteilungen, d. h. der Klassenanzahl, fällt dieses bei den Items SKF12 und SA75 im Vergleich

zu dem linksschief verteilten Item SKV55, welches hinsichtlich seiner Kennwerte als repräsentativ für die restliche Itemstichprobe angesehen werden kann, deutlich höher aus ( $\text{Verhältnis}_{\text{SKF12}} = .18$ ,  $\text{Verhältnis}_{\text{SA75}} = .23$ ,  $\text{Verhältnis}_{\text{SKV55}} = .03$ ). Zieht man das Kriterium der Itemverteilung heran, wird der Ausschluss dieser beiden Items aufgrund der starken Variation über die Klassen hinweg empfohlen.

Zieht man ergänzend das Kriterium des Einsatzzwecks des Fragebogens in der Unterrichtsentwicklung heran, lassen sich keine zusätzlichen Eliminationsentscheidungen ableiten. Linksschiefe Verteilungen erfüllen aus motivationstheoretischen Aspekten heraus einen wichtigen Zweck für die Lehrkräfte in der Unterrichtsentwicklung. Derartige Antwortverteilungen motivieren die Lehrkraft aufgrund der überwiegend positiven Einschätzung. Auch rechtsschiefe Items sind in der Unterrichtsentwicklung hilfreich, denn diese zeigen Entwicklungspotentiale der Lehrkräfte. Hinsichtlich gleichverteilter Items bleibt offen, inwiefern die Motivationsmöglichkeit, als Lehrkraft durch eine entsprechende Verhaltensänderung den Ausschlag in die positive Richtung zu geben, die Tatsache ausgleicht, dass kein eindeutiges Ergebnismuster hinsichtlich Zustimmung bzw. Ablehnung vorliegt und dieses Item bei einer Einpunktmessung daher keinen Informationsgewinn im statistischen Sinn liefert.

#### 9.1.4 Analyse der Lokationsmaße

Die Items weisen einen *durchschnittlichen Mittelwert* von  $M = 2.90$  auf (ich-Version:  $M = 2.89$ ; wir-Version:  $M = 2.90$ ), d. h. den Items zur Unterrichtswahrnehmung wird im Durchschnitt eher zugestimmt. Werden die Itemmittelwerte in eine *Rangreihe* gebracht, zeigt sich, dass das Item SLK22 mit einem Mittelwert von  $M_{\text{insgesamt}} = \text{ich-Version} = 3.67$  den *höchsten Mittelwert* in der Gesamtstichprobe und der ich-Version aufweist. In der wir-Version hat das Item SKV46 mit einem Mittelwert von  $M = 3.68$  den höchsten Zustimmungsindeks, darauf folgt das Item SLK22 mit einem Mittelwert von  $M = 3.67$ . Danach folgen mit einem Mittelwert von mindestens  $M = 3.50$  in der Gesamtstichprobe und mit geringfügigen Abweichungen auch in der ich- und wir-Version die Items: SKV46 ( $M_{\text{insgesamt}} = 3.65$ ;  $M_{\text{ich-Version}} = 3.63$ ), SLK18 ( $M_{\text{insgesamt}} = 3.60$ ;  $M_{\text{ich-Version}} = 3.62$ ;  $M_{\text{wir-Version}} = 3.57$ ), SLK23 ( $M_{\text{insgesamt}} = 3.53$ ;  $M_{\text{ich-Version}} = 3.53$ ;  $M_{\text{wir-Version}} = 3.54$ ), SLK30 ( $M_{\text{insgesamt}} = 3.52$ ;  $M_{\text{ich-Version}} = 3.49$ ;  $M_{\text{wir-Version}} = 3.55$ ), wobei in der ich-Version auch noch das Item SKF15 einen Mittelwert von  $M = 3.51$  aufweist. Somit weisen v. a. Items zur Erfassung des Lernförder-

lichen Klimas hohe Zustimmungswerte auf, die sich auf die Freundlichkeit der Lehrkraft im Umgang mit den Schülern beziehen.

Der *niedrigste Mittelwert* zeigt sich in der Gesamtstichprobe und der wir-Version bei dem Item SA77 mit einem Itemmittelwert von  $M_{\text{insgesamt}} = 1.80$  ( $M_{\text{wir-Version}} = 1.82$ ). In der ich-Version ist es das Item SA72 mit einem Mittelwert von  $M = 1.65$ . Die in der Rangreihe folgenden vier Items sind: SA72 ( $M_{\text{insgesamt}} = 1.81$ ;  $M_{\text{wir-Version}} = 1.97$ ), SA69 ( $M_{\text{insgesamt}} = 1.88$ ;  $M_{\text{ich-Version}} = 1.72$ ;  $M_{\text{wir-Version}} = 2.04$ ), SA70 ( $M_{\text{insgesamt}} = 2.01$ ;  $M_{\text{ich-Version}} = 1.94$ ;  $M_{\text{wir-Version}} = 2.08$ ), SA71 ( $M_{\text{insgesamt}} = 2.12$ ;  $M_{\text{ich-Version}} = 2.07$ ;  $M_{\text{wir-Version}} = 2.17$ ). Folglich stimmt der Großteil der Schüler bei Aktivierungsitems und insbesondere bei Items nicht zu, die erfassen, ob die Schüler Aufgaben selbst erfinden durften, ob die Schüler die Lehrkraft durch die Bedienung technischer Geräte unterstützt haben, etwas am Overhead-Projektor dargestellt haben, sich gegenseitig aufrufen durften oder von sich aus Vorschläge (z. B. zum Ablauf der Unterrichtsstunde) gemacht haben. Wird der Medianwert vergleichend hinzugezogen, zeigt sich, dass die Tendenzaussagen aufgrund der Analyse des Medianwertes den auf der Basis des Mittelwerts gefundenen Aussagen bezüglich der Extrema nicht widersprechen. Die Zugrundlegung des Medians als Zustimmungssindex würde die Itemanzahl bei jeder Kategorie aufgrund der Informationsreduktion dieses Maßes lediglich vergrößern. So würden z. B. insgesamt 18 Items in der Gesamtstichprobe (ich-Version: 16 Items; wir-Version: 18 Items) einen Median von 4 aufweisen und somit alle zusammen das positive Extrem bilden.

In Abbildung 12 ist die Verteilung der Itemmittelwerte in der Gesamtstichprobe innerhalb der theoretisch postulierten *Unterrichtsmerkmale* für den Mittelwert als Zustimmungssindex dargestellt. Es zeigt sich deskriptiv, dass die Zustimmung zu den Aktivierungsitems entsprechend obiger Befunde im Durchschnitt am geringsten ( $M_A = 2.50$ ) ist. Die durchschnittliche Schülerzustimmung zu den Motivierungsindikatoren liegt mit einem Mittelwert von  $M_M = 2.67$  knapp oberhalb des theoretischen Mittelwerts der Antwortskala. Der durchschnittliche Itemmittelwert der anderen vier Unterrichtsmerkmale liegt mindestens eine halbe Antwortkategorie oberhalb des Skalenmittelwerts und ist bei dem Unterrichtsmerkmal Lernförderliches Klima mit  $M_{LK} = 3.31$  deskriptiv am höchsten.

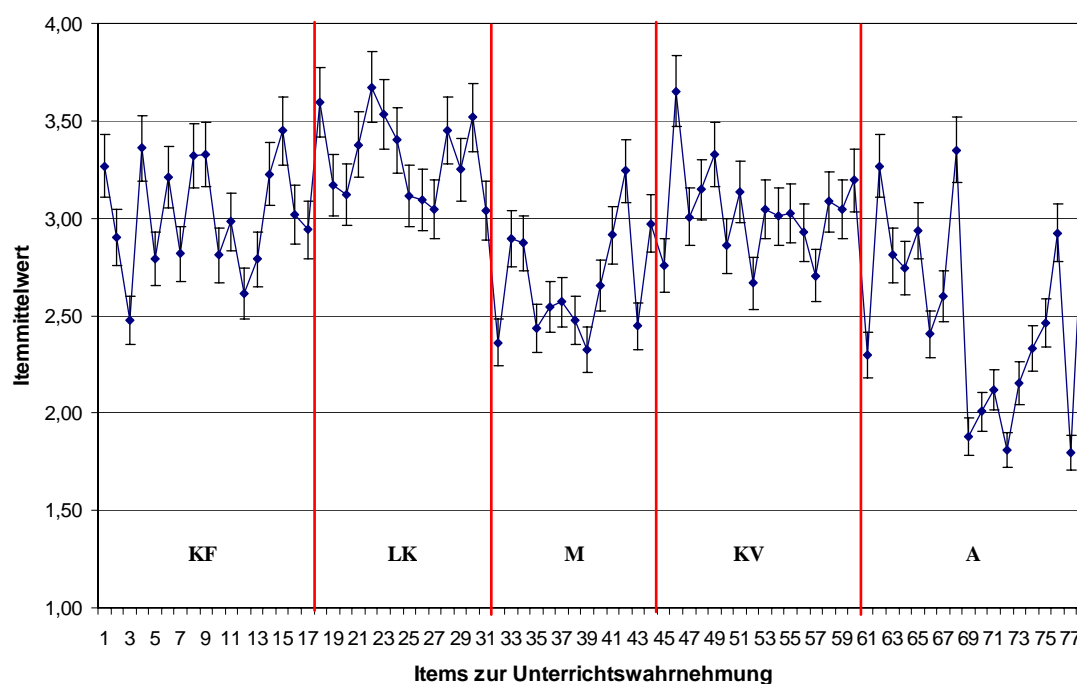


Abbildung 12: Verteilung der Itemmittelwerte zur Unterrichtswahrnehmung innerhalb der theoretisch postulierten Unterrichtsmerkmale (KF = Klassenführung, LK = Lernförderliches Klima, M = Motivierung, KV = Klarheit und Verständlichkeit, A = Aktivierung) mit Angabe der 95%-Konfidenzintervalle

Die größte Spannweite der Itemmittelwerte liegt in der Gesamtstichprobe bei den Aktivierungsitems mit 1.55 als Differenz zwischen  $M_{\text{MAX}}$  (SA68) = 3.35 und  $M_{\text{MIN}}$  (SA77) = 1.80 vor (zum Vergleich:  $\text{Spannweite}_{\text{KV}} = 0.99$ ,  $\text{Spannweite}_{\text{KF}} = 0.97$ ,  $\text{Spannweite}_{\text{M}} = 0.92$  und  $\text{Spannweite}_{\text{LK}} = 0.64$ ).

Werden alle Schülerangaben der *ich-* und *wir-Version* verglichen, zeigt sich eine durchschnittliche Abweichung der Itemmittelwerte insofern, dass der Itemmittelwert der *ich-Version* bei 36 Items geringer ist als der Itemmittelwert der *wir-Version* und bei ebenfalls 36 Items trifft der umgekehrte Fall zu. Sechs Items (SM34, SLK22, SLK31, SK17, SKV55, SLK21) weisen den gleichen Zustimmungssindex in der *ich-* und *wir-Version* auf. Die größte positive Differenz zwischen dem Itemmittelwert der *ich-* und dem der *wir-Version* weist das Item SA78 („Ich bin / Alle Schüler sind während der Unterrichtsstunde mindestens einmal drangekommen“) auf ( $M_{\text{ich-Version}} = 3.41$ ;  $M_{\text{wir-Version}} = 2.61$ ). Die größte negative Abweichung zeigt das Item SA74 („Ich habe / Wir haben Lösungswege gesucht und sie der Klasse vorgestellt“) ( $M_{\text{ich-Version}} = 2.09$ ;  $M_{\text{wir-Version}} = 2.58$ ). Dies entspricht in beiden Fällen den Erwartungen, denn einzelne Schüler werden in jedem Fall, aber nicht unbedingt immer alle von der Lehrkraft aufgerufen (SA78) und i.d.R. wird in der Klasse

gemeinsam nach verschiedenen Lösungswegen gesucht, der Einzelne hingegen bekommt selten direkt einen solchen Arbeitsauftrag (SA74).

Zieht man die *Unterrichtsmerkmale* als Vergleichskriterium heran, wie in Abbildung 13 dargestellt, zeigt sich, dass das Merkmal *Aktivierung* die größten Unterschiede zwischen der ich- und wir-Version aufweist. Die Summe der Beträge positiver und negativer Abweichungen im Itemmittelwert zwischen der ich- und wir-Version beträgt hier 3.71 und ist im Vergleich mit den Abweichungssummen der anderen Merkmalsbereiche der mit Abstand größte Wert:  $\text{Summe}_{\text{KV}} = 1.58$ ,  $\text{Summe}_{\text{KF}} = 0.86$ ,  $\text{Summe}_{\text{LK}} = 0.84$  und  $\text{Summe}_{\text{M}} = 0.79$ .



Abbildung 13: Verteilung der Itemmittelwerte in Abhängigkeit von der Fragebogenversion sowie des Unterrichtsmerkmals (KF = Klassenführung, LK = Lernförderliches Klima, M = Motivierung, KV = Klarheit und Verständlichkeit, A = Aktivierung)

Ferner beträgt die Relation der Anzahl der Items, die eine Abweichung von  $\leq 0.05$  zum Itemmittelwert der wir-Version aufweisen 0.17, da nur drei von insgesamt 18 Items dieses Kriterium erfüllen. Zum Vergleich die Verhältnisswerte der anderen Merkmalsbereiche:

$$\text{Relation}_{\text{KF}} = \frac{13}{17} = .76, \text{Relation}_{\text{LK}} = \frac{9}{14} = .64, \text{Relation}_{\text{M}} = \frac{8}{13} = .62, \text{Relation}_{\text{KV}} = \frac{9}{16} =$$

.56. Dieser Vergleich zeigt, dass auch hier das Merkmal *Aktivierung* den größten Anteil an Items enthält, deren Differenz zum Itemmittelwert der wir-Version größer 0.05 ist. Dabei überwiegen die negativen Abweichungen, d. h. die Schüler stimmen den Items in der wir-Version im Durchschnitt um 0.18 Mittelwertseinheiten bei einem Streubereich von [0.03; 0.49] stärker zu.

Betrachtet man ausschließlich die Abweichungssumme, weist das Merkmal *Motivierung* insgesamt mit einer Abweichungssumme von 0.79 und einer durchschnittlichen Abweichung von 0.03 bei einem Streubereich von [-0.20; 0.13] die geringsten Abweichungen

zwischen der ich- und wir-Version auf. Diese sind bis auf das Item SM42 ( $d_{\text{ich-Version} - \text{wir-Version}} = -0.20$ ) positiv, d. h. in der ich-Version ist die Zustimmung zu diesen Items tendenziell höher. Wird allerdings die Relation zur gesamten Itemanzahl pro Unterrichtsmerkmal mit berücksichtigt, so weisen die Items des Unterrichtsmerkmals Klassenführung die größte Relation und somit den größten Anteil an Items auf, deren Unterschied zwischen dem Wert der ich- und der wir-Version  $\leq 0.05$  ist.

Werden die 25 *neutralen Items* betrachtet, die sowohl in der ich- als auch der wir-Version gleich formuliert sind, so zeigt sich eine durchschnittliche Abweichung des Itemmittelwerts der ich-Version von der wir-Version von  $M_{\text{insgesamt}} = 0.02$  mit einem Streubereich von  $[-0.03; 0.11]$ . Dabei beträgt die durchschnittliche positive Abweichung der ich-Version von der wir-Version bei 14 Items im Mittel  $M_{\text{positive Differenz}} = 0.04$  mit einem Streubereich von  $[0.01; 0.11]$ . Bei acht Items zeigt sich eine negative Abweichung mit einem Mittelwert von  $M_{\text{negative Differenz}} = -0.02$  bei einem Streubereich von  $[-0.03; -0.01]$ . Die drei Items SKF17, SLK22 und SLK31 stimmen in ihrem Itemmittelwert in beiden Fragebogenversionen überein. Das Item SLK26 („Der Lehrer machte während des Unterrichts ab und zu Scherze“) weicht mit 0.11 am stärksten in die positive Abweichung ab und SM32 („Der Lehrer hat Verbindungen zu anderen Fächern hergestellt“) mit 0.10 am zweitstärksten. Die größten aber absolut gesehen sehr geringen negativen Abweichungen weisen mit -0.03 die Items SKF3, SKF6 und SA75 auf.

Von den 53 *ich- bzw. wir-bezogenen Items* haben insgesamt 21 Items im Durchschnitt einen um 0.13  $[0.01; 0.80]$  höheren Itemmittelwert in der ich-bezogenen Formulierung als in der wir-bezogenen Formulierung auf. Am größten erweist sich die Abweichung bei Item SA78. Wie bereits zu Beginn dieses Abschnitts berichtet, beträgt der Mittelwert der Items, die ich-bezogen formuliert sind  $M_{\text{ich}} = 2.89$  und der der Items, die wir-bezogen formuliert sind  $M_{\text{wir}} = 2.90$ . Die drei Items SLK21, SM34 und SKV55 stimmen hinsichtlich des Mittelwerts überein. 29 Items weisen in der ich-bezogenen Formulierung einen im Durchschnitt um 0.15 geringeren Mittelwert bei einem Streubereich von  $[-0.49; -0.04]$  auf. Die größte negative Abweichung ergibt sich dabei für das Item SA74 mit einem Mittelwertsunterschied von 0.49 ( $M_{\text{ich}} = 2.09$ ,  $M_{\text{wir}} = 2.58$ ).

Wird der *theoretische Mittelwert* der vierstufigen Antwortskala mit  $M = 2.50$  als absolutes Vergleichskriterium angesetzt, ergibt sich folgendes Bild: 62 Items (ich-Version: 61 Items, wir-Version: 63 Items) weisen in der Gesamtstichprobe eine durchschnittliche positive

Differenz  $d_{\text{Itemmittelwert} - \text{theoretischer Mittelwert}}$  von 0.57 (ich-Version:  $M_d = 0.58$ , wir-Version:  $M_d = 0.56$ ) zu diesem auf. Die Itemmittelwerte liegen somit im Durchschnitt ca. eine halbe Antwortkategorie oberhalb des theoretischen Mittelwerts. Dies geht in diesem Kontext mit einer höheren Ausprägung der Unterrichtsmerkmalsindikatoren einher. Dabei weichen die Items SA67 („Wir haben / Ich habe in der Stunde eigene Arbeitsergebnisse vorgestellt“) und das bereits angeführte Item SA74 in der ich- und wir-Formulierung derart ab, dass die Zustimmung in der wir-bezogenen Formulierung oberhalb, in der ich-bezogenen Formulierung jedoch unterhalb des theoretischen Skalenmittels liegt. Dies könnte erneut darauf zurückzuführen sein, dass Schüler entsprechende Anweisungen eher im Klassenkontext erhalten und entsprechend den qualitativen Voruntersuchungen die Beschreibung ‚eigene Arbeitsergebnisse‘ mit einem höheren Anspruch an den Leistungsumfang in der ich-Formulierung assoziieren als in der wir-Formulierung. Die durchschnittliche negative Abweichung der restlichen 16 Items (ich-Version: 17 Items, wir-Version: 15 Items) vom theoretischen Mittelwert beträgt  $M_d = -0.26$  (ich-Version:  $M_d = -0.30$ ,  $M_d = -0.23$ ). Die Tatsache, dass die Itemmittelwerte tendenziell eher oberhalb der Skalenmitte liegen und diese Abweichung in die positive Richtung durchschnittlich größer ist als die Abweichungen unterhalb der Skalenmitte bedingt die überwiegend vorgefundene Linksschiefe der Itemverteilungen entscheidend mit.

Hinsichtlich der *Unterrichtsmerkmale* liegen in der Gesamtstichprobe und mit einer Ausnahme auch in der ich- und wir-Version v. a. Items mehr als eine Antwortkategorie über dem theoretischen Mittel, die das Unterrichtsmerkmal *Lernförderliches Klima* erfassen: SLK22 ( $d = 1.17$ ), SKV46 ( $d = 1.15$ ), SLK18 ( $d = 1.10$ ), SLK23 ( $d = 1.03$ ) und SLK30 ( $d = 1.02$ , ich-Version:  $d = 0.99$ ). In der ich-Version weist noch das Item SKF15 zusätzlich eine Differenz  $d$  von 1.01 auf.

Die größten negativen Abweichungen von der theoretischen Skalenmitte weisen in der Gesamtstichprobe durchgängig Items zur Erfassung des Unterrichtsmerkmals *Aktivierung* auf: SA77 ( $d = -0.70$ ), SA72 ( $d = -0.69$ ), SA69 ( $d = -0.62$ ), SA70 ( $d = -0.49$ ), SA71 ( $d = -0.38$ ), SA73 ( $d = -0.35$ ) und SA61 ( $d = -0.20$ ). In der ich- und wir-Version zeigt sich eine ähnliche Rangfolge, wobei die Differenzwerte bei der ich-Version durchgängig höher sind als bei der wir-Version:

- SA73 ( $d_{\text{ich-Version}} = -0.85$ ,  $d_{\text{wir-Version}} = -0.53$ )
- SA69 ( $d_{\text{ich-Version}} = -0.78$ ,  $d_{\text{wir-Version}} = -0.46$ )

- SA77 ( $d_{\text{ich-Version}} = -0.72$ ,  $d_{\text{wir-Version}} = -0.68$ )
- SA70 ( $d_{\text{ich-Version}} = -0.56$ ,  $d_{\text{wir-Version}} = -0.42$ )
- SA71 ( $d_{\text{ich-Version}} = -0.43$ ,  $d_{\text{wir-Version}} = -0.33$ )

Diese Tendenz, dass die Abweichung vom theoretischen Mittelwert bei der ich-Version größer ist als bei der wir-Version setzt sich bei weiteren 42 Items fort. Dies zeigt, dass die Schülerangaben in der ich-Version extremer bzw. in der wir-Version weniger extrem ausfallen, was auf die Effekte der kognitiven Aggregation über alle Schüler in den wir-bezogenen Items zurückgeführt werden könnte. Innerhalb der Unterrichtsmerkmale zeigt sich, dass die Merkmale Klassenführung (bis auf das Item SKF3 mit einem Itemmittelwert von  $M_{\text{SKF3}} = 2.47$ ), Lernförderliches Klima und Klarheit und Verständlichkeit ausschließlich positive Abweichungen vom theoretischen Mittelwert aufweisen, d. h. durchgängig größer als  $M = 2.50$  sind. Wohingegen die Merkmale Motivierung und Aktivierung ein heterogenes Bild aufzeigen mit einem Teil der Items, die oberhalb und einem anderen Teil der Items, die unterhalb des theoretischen Mittelwerts liegen.

### 9.1.5 Analyse der Dispersionsmaße

Für alle Items ergibt sich insgesamt aber auch bei separater Betrachtung der ich- und wir-Version jeweils als *Minimum* der Wert 1 und als *Maximum* der Wert 4. Somit wird die gesamte Breite der Antwortkategorien genutzt. Die Verteilungen der Schülerangaben liegen hinsichtlich der relativen Häufigkeiten der Antwortkategorien in der Gesamtstichprobe zwischen 1.71% (ich-Version: 1.69%; wir-Version: 1.73%) bei der ersten Antwortkategorie des Items SKV46 und 76.56% (ich-Version: 75.65%, wir-Version: 77.48%) bei der vierten Antwortkategorie des Items SLK22. Die Antworten liegen über alle Probanden hinweg sowie ebenfalls in der ich- und wir-Version eher in der rechten Skalenhälfte, d. h. der Anteil der Zustimmungen ist größer als der der Ablehnung des Items, wie Tabelle 13 entnommen werden kann.

Tabelle 13: *Relative mittlere Häufigkeit pro Antwortkategorie in der Gesamtstichprobe sowie den Teilstichproben der ich- und wir-Version.*

Fragebogen- version	Antwortkategorien			
	1	2	3	4
insgesamt	15.60%	17.21%	29.19%	38.00%
ich-Version	15.85%	17.33%	28.98%	37.84%
wir-Version	15.33%	17.09%	29.41%	38.17%



Die Items zur Unterrichtswahrnehmung streuen relativ stark, d. h. die Schüler unterscheiden sich sehr in der Beantwortung der Items. Die *mittlere Standardabweichung* liegt bei  $sd_{\text{insgesamt}} = sd_{\text{ich-Version}} = sd_{\text{wir-Version}} = 0.98$ . 40 Items weisen in der Gesamtstichprobe und der ich-Version mindestens eine Standardabweichungseinheit auf, in der wir-Version sind es 39 Items. Bei den anderen zeigt sich eine Standardabweichung kleiner 1.0. Das Minimum des Standardabweichungskennwerts beträgt 0.66 beim Item SKF46 (ich-Version:  $SD_{\text{MIN}}(\text{SLK22}) = 0.66$ , wir-Version:  $SD_{\text{MIN}}(\text{SKV46}) = 0.65$ ) und das Maximum 1.22 beim Item SKF12 (ich-Version:  $SD_{\text{MAX}}(\text{SA75}) = 1.22$ , wir-Version:  $SD_{\text{MAX}}(\text{SKF12}) = 1.24$ ).

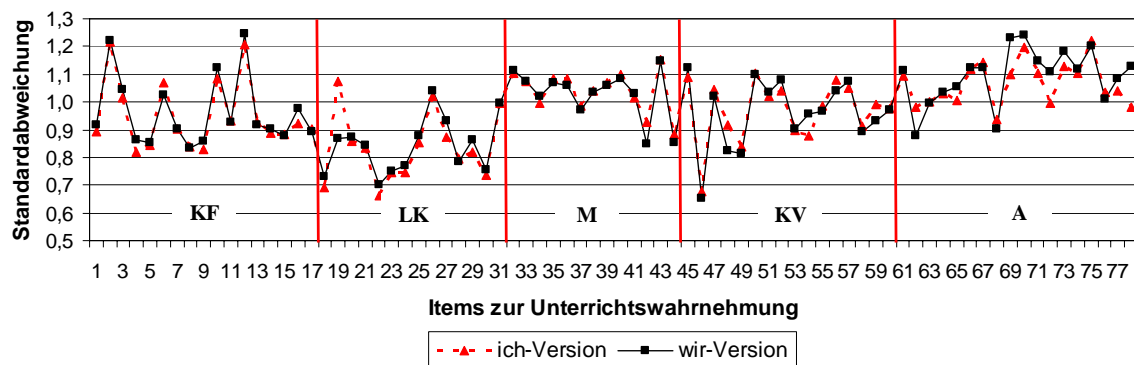


Abbildung 14: Darstellung der Itemstandardabweichungen zur Unterrichtswahrnehmung in Abhängigkeit von der Fragebogenversion sowie des Unterrichtsmerkmals (KF = Klassenführung, LK = Lernförderliches Klima, M = Motivierung, KV = Klarheit und Verständlichkeit, A = Aktivierung)

Abbildung 14 stellt die Standardabweichung der Items zur Unterrichtswahrnehmung in Abhängigkeit von der Frageformulierung und dem Unterrichtsmerkmal dar. Es wird deutlich, dass die Standardabweichungen der Items zur Erfassung des Merkmals *Lernförderliches Klima* im Durchschnitt am geringsten sind ( $\overline{SD}_{\text{insgesamt}} = \overline{SD}_{\text{wir-Version}} = \overline{SD}_{\text{ich-Version}} = 0.84$ ), d. h. die Schüler stimmen insgesamt bei der Beantwortung rein deskriptiv stärker überein als bei den anderen Merkmalsindikatoren. Es ist auch ersichtlich, dass die *Aktivierungsitems* im Durchschnitt die höchsten Standardabweichungswerte aufweisen ( $\overline{SD}_{\text{insgesamt}} = \overline{SD}_{\text{wir-Version}} = 1.09$ ,  $\overline{SD}_{\text{ich-Version}} = 1.07$ ). Die Spannweite der Itemstandardabweichungen ist bei den Items zur Erfassung des Merkmals *Klarheit und Verständlichkeit* mit einer Differenz von  $SD_{\text{MAX}}(\text{SKV45}) - SD_{\text{MIN}}(\text{SKV46}) = 0.44$  am größten (zum Vergleich:  $\text{Spannweite}_{\text{KF}} = 0.39$ ,  $\text{Spannweite}_{\text{LK}} = 0.35$ ,  $\text{Spannweite}_{\text{A}} = 0.30$  und  $\text{Spannweite}_{\text{M}} = 0.28$ ). Werden die Itemstandardabweichungen der *ich- und wir-Version* miteinander verglichen, zeigen sich tendenziell eher geringe Abweichungen. Bei 40 Items ist der Standardabwei-

chungswert in der ich-Version um durchschnittlich 0.04 geringer. Die Standardabweichung stimmt bei 8 Items überein und 30 Items weisen eine im Durchschnitt um 0.03 größere Standardabweichung in der ich-Version auf. Die größte positive Differenz zwischen der Standardabweichung der ich- und wir-Version findet sich mit der Differenz von 0.21 bei dem Item SLK19 ( $SD_{\text{ich-Version}} = 1.07$ ,  $SD_{\text{wir-Version}} = 0.87$ ) und die größte negative Abweichung der ich-Version von der wir-Version weist mit -0.15 das Item SA78 ( $SD_{\text{ich-Version}} = 0.98$ ,  $SD_{\text{wir-Version}} = 1.13$ ) auf.

Die größten Unterschiede zwischen der ich- und wir-Version weisen die Aktivierungsitems mit einer *Abweichungssumme* von 0.86 über die Beträge der einzelnen positiven und negativen Abweichungen in der Standardabweichung hinweg auf und die kleinste Abweichungssumme haben die Motivierungsitems mit 0.24 (zum Vergleich:  $\text{Summe}_{\text{KF}} = 0.36$ ,  $\text{Summe}_{\text{LK}} = \text{Summe}_{\text{KV}} = 0.51$ ). Wird die gesamte Itemanzahl über den Relationswert der Items, die eine Standardabweichungsdifferenz von  $\leq 0.05$  zwischen der ich- und wir-Version aufweisen zur Gesamtitemanzahl berücksichtigt, so weisen alle Klassenführungsitems eine Abweichung zwischen der ich- und wir-Version von  $\leq 0.05$  auf. Die  $\text{Relation}_{\text{KF}}$  beträgt somit  $\frac{17}{17} = 1$ . Die kleinste Relation und somit den größten Anteil an Items über 0.05

Standardabweichungseinheiten haben Aktivierungsitems mit einer Relation von  $\text{Relation}_{\text{A}} = \frac{14}{18} = .64$  (zum Vergleich:  $\text{Relation}_{\text{M}} = \frac{12}{13} = .92$ ,  $\text{Relation}_{\text{KV}} = \frac{13}{16} = .81$  und  $\text{Relation}_{\text{LK}} = \frac{12}{14} = .86$ ).

Die *neutralen Items* zeigen eine durchschnittliche Abweichung der Standardabweichung in der ich- und wir-Version von -0.01 [-0.05; 0.04] auf. Die drei Items SKF11, SLK31 und SM33 stimmen überein. 15 Items weisen eine geringe durchschnittliche negative Abweichung von -0.05 auf und sieben eine ebenfalls geringe positive mittlere Abweichung von 0.01. Betrachtet man ausschließlich die 53 Items, die *ich-* bzw. *wir-bezogen* formuliert sind und lässt neutral formulierte Items außen vor, dann weisen 25 Items eine negative Differenz zwischen den beobachteten Werten der ich- und wir-Version zwischen -0.01 und -0.15 auf, die bedeutet, dass die Unterrichtswahrnehmungen bei der Beurteilung in der ich-Version weniger variieren und somit eine geringere Standardabweichung aufweisen als bei der wir-Version. Fünf Items differieren nicht und 23 Items weisen eine positive Differenz zwischen 0.01 und 0.21 auf, d. h. die Angaben in der ich-Version variieren stärker. Die Differenzwer-

te sind dabei allerdings recht gering und liegen zwischen -0.01 und -0.15 bzw. zwischen 0.01 und 0.21. Dabei weist das Item SLK19 die größte positive Abweichung mit 0.21 Standardabweichungseinheiten und das Item SA78 mit -0.15 die größte negative Abweichung auf.

### **9.1.6 Zusammenfassende Bewertung**

Die deskriptivstatistischen Analysen zeigen, dass die Schüler den Itemaussagen zur Unterrichtswahrnehmung tendenziell eher zustimmen. Dies zeigt sich in der überwiegenden Linksschiefe der Items und einem Gesamtmittelwert, der oberhalb des theoretischen Skalenmittels liegt. Dabei wird allerdings die gesamte Breite der Antwortkategorien ausgeschöpft. In Bezug auf die Itemselektion stechen folgende Items aufgrund auffälliger Analyseergebnisse hervor:

- Die Items SKF10 („Bestimmte Aufgaben sind von dazu eingeteilten Schülern/innen (z. B. Klassendienste) erledigt worden“) und SM36 („Die Arbeitsblätter haben mir/uns Spaß gemacht“) aufgrund einer relativ hohen Missinganzahl.
- Die Items SKF12 („Einzelne Schüler/innen haben dem Lehrer geholfen (z. B. beim Austeilen von Arbeitsblättern, dem Abwischen der Tafel oder anderen Dingen“) und SA75 („Der Lernstoff wurde durch Abbildungen veranschaulicht (z. B. Mindmap)“ wegen ihrer zweigipfligen Verteilungsform.

Im Folgenden werden die Items zur Unterrichtswahrnehmung ausführlichen Item- und Reliabilitätsanalysen unterzogen.

## **9.2 Item- und Reliabilitätsanalysen**

Drei Kennwerte werden in diesem Abschnitt untersucht: die Trennschärfe, der Zustimmungsindeks als Äquivalent für das klassische Maß der Schwierigkeit und Homogenitätsmaße. Diese werden zunächst im Abschnitt 9.2.1 näher erläutert. Eine Übersicht zu den Kennwerten auf Itemebene findet sich im Anhang D.2.1. Da es sich bei dem Fragebogen in Bezug auf die Unterrichtsmerkmale um ein mehrdimensionales Evaluationsinstrument handelt, werden die Itemanalysen separat für die jeweiligen postulierten Subskalen durchgeführt (Abschnitt 9.2.2 bis 9.2.6). Im Abschnitt 9.2.7 werden die Skalen im Hinblick auf eine Gesamtbetrachtung des Fragebogens bewertet.

### 9.2.1 Auswertungsvorgehen

Die *Trennschärfe* ist ein Maß für die Prototypikalität eines Items hinsichtlich der jeweilig zugehörigen Skala. Bei intervallskalierten Ratingskalen wird die Trennschärfe aus der Produkt-Moment-Korrelation zwischen der Itemangabe und dem Summenwert einer Skala berechnet (Amelang & Schmidt-Atzert, 2006). Zur Interpretation wird die part-whole-korrigierte Trennschärfe  $r_{itc}$  herangezogen, bei der das betreffende Item nicht mit in den

Skalenwert eingeht (Bühner, 2006):  $r_{itc} = \frac{r_{it} sd_t - sd_i}{\sqrt{sd_t^2 + sd_i^2 - 2 r_{it} sd_t sd_i}}$ , wobei  $r_{it}$  die Korrela-

tion zwischen dem Item  $i$  und der Skala  $t$  und  $sd$  die jeweilige Standardabweichung kennzeichnet. Gemäß Fisseni (1997) werden Trennschärfewerte von  $r_{itc} < .30$  als niedrig, eine Trennschärfe von  $.30 \leq r_{itc} \leq .50$  als mittel und  $r_{itc} > .50$  als hoch eingestuft. Erfolgt die Itemselektion nur anhand der Trennschärfe, besteht die Gefahr, Items mit extremen Zustimmungswerten zu eliminieren. Daher wird nach Lienert und Raatz (1998) zusätzlich der *Selektionskennwert*  $SK$  bestimmt, der auf der Basis der korrigierten Trennschärfewerte  $r_{itc}$  und der Standardabweichung  $sd$  jene Items mit extremen Zustimmungswerten nach

oben korrigiert:  $SK = \frac{r_{itc}}{2 sd}$ .

Der Schwierigkeitsindex, bzw. im Kontext der Unterrichtswahrnehmung ist der Terminus *Zustimmungswert* sinnvoller, wird hinsichtlich seines Einflusses auf das Trennschärfemaß und die Homogenitätskennwerte betrachtet. Die Wahrscheinlichkeit einer hohen Trennschärfe ist für Items mit einem mittleren Zustimmungswert größer, was wiederum eine höhere Reliabilität für die jeweilige Skala erwarten lässt. Bei einer mehrstufigen Antwortskala und dem Vorliegen des Intervallskalenniveaus ist es üblich, den Itemmittelwert als Äquivalent für den Zustimmungswert heranzuziehen (Bühner, 2006; Amelang & Schmidt-Atzert, 2006). Um die Vergleichbarkeit hinsichtlich der Skalierung mit dem klassischen Schwierigkeitsindex zu gewährleisten, wird dem Vorschlag von Fisseni (1997) folgend ein

mehrstufiger Zustimmungswert  $P_m$  berechnet:  $P_m = \frac{\sum x_{emp}}{\sum x_{max}}$ , wobei  $\sum x_{emp}$  dem beobachte-

ten Summenwert des jeweiligen Items und  $\sum x_{max}$  dem maximalen Summenwert pro Item entspricht. Da die Analyse des Zustimmungswertes ohne die gleichzeitige Betrachtung des Streuungsmaßes wenig aussagekräftig ist, wird dieses mit Hilfe von graphischen Verfahren in die Analyse einbezogen.

Des Weiteren gilt es, die *Homogenität* zu untersuchen, die den Grad angibt, in dem die Items eines Tests dasselbe Konstrukt messen (Fischer, 1974). Als Maße der klassischen Testtheorie kommen die Interkorrelation der Items und faktorenanalytische Verfahren in Frage (Fisseni, 1997; Rost, 1996). Die Ergebnisse der exploratorischen und konfirmatorischen Faktorenanalysen werden im Abschnitt 9.3 aufgezeigt. Hinsichtlich der Interkorrelation als Homogenitätsmaß gibt der Kennwert *MIC* die mittlere Interitemkorrelation einer Skala an. Gemäß Bühner (2006) liegt der MIC-Wert in der Regel zwischen .20 und .40. Bei postulierter Eindimensionalität der Skalenitems sollte der MIC-Wert möglichst hoch sein. Allerdings ist auch dieser Wert problematisch, wenn manche Items hoch und manche niedrig miteinander korrelieren. Eine perfekte Interkorrelation kann außerdem nur bei gleichen Itemmittelwerten erreicht werden. Folglich bedingen auch die Mittelwertsunterschiede die Interkorrelationshöhe und nicht nur die inhaltliche Homogenität (Lienert & Raatz, 1998). Deshalb schlägt Cortina (1993) vor, die *Präzision von Alpha*  $P_\alpha$  als einen weiteren Indikator für die Homogenität der Items zu betrachten. Dieser wird folgendermaßen berechnet:

$$P_\alpha = \frac{sd_{MIC}}{\sqrt{(0,5 c (c - 1)) - 1}}, \text{ wobei } sd_{MIC} \text{ die Standardabweichung der}$$

mittleren Interitemkorrelation angibt und  $c$  die Itemanzahl. Bereits kleine Werte von  $P_\alpha = .01$  können darauf hinweisen, dass die analysierte Skala nicht eindimensional ist (Cortina, 1993).

*Cronbachs  $\alpha$*  als Maß der internen Konsistenz wird ebenfalls hinzugezogen. Dieses Maß wird gemäß Cronbach (1951, zitiert nach Bühner, 2006) anhand folgender Formel berechnet:

$$\text{net: } \text{Cronbachs } \alpha = \frac{c}{c - 1} \left( 1 - \frac{\sum_{i=1}^j sd_i^2}{sd_t^2} \right), \text{ wobei } sd_i^2 \text{ die Itemvarianz, } c \text{ die Anzahl der Items}$$

und  $sd_t^2$  die Varianz des Skalengesamtwertes erfasst. Cronbachs  $\alpha$  ist somit eine Funktion der Summe der Itemvarianzen in Relation zur Gesamtvarianz. Mit steigender Itemanzahl nimmt auch Cronbachs  $\alpha$  bei positiv korrelierten Items zu (Bühner, 2006). Gemäß Fisseni (1997) kann Cronbachs  $\alpha$  bei einem Wert von  $r_{tt} < .80$  als niedrig, zwischen  $.80 \leq r_{tt} \leq .90$  als mittel und ab  $r_{tt} > .90$  als hoch eingestuft werden. Bühner (2006) gibt an, dass bei vorliegenden stark unterschiedlichen Itemvarianzen das *standardisierte Cronbachs  $\alpha$*  verwendet werden sollte, welches auf den Korrelationen z-standardisierter Items beruht. Die vor-

liegenden Daten zeigen, dass die Itemvarianzen mit einem Streubereich von [0.44; 1.50] in der Gesamtstichprobe, einem von [0.44;1.49] in der ich-Version sowie einem von [0.42; 1.55] in der wir-Version relativ stark streuen und somit das standardisierte Cronbachs  $\alpha$  hinsichtlich der Interpretation verwendet werden sollte. Im Folgenden werden die beschriebenen Kennwerte separat für die einzelnen Unterrichtsmerkmale ermittelt.

### 9.2.2 Skala Klassenführung

Dieser Skalenanalyse liegen in der Gesamtstichprobe 1286 (ich-Version: 642, wir-Version: 644) Schülerangaben zugrunde. Cronbachs  $\alpha$  basierend auf 17 Items ergibt einen Koeffizienten von .814 (ich-Version: .818, wir-Version: .811) und liegt gemäß Fisseni (1997) im mittleren Bereich. Die Präzision von Cronbachs  $\alpha$  mit  $P_\alpha = 0.01$  ist gerade noch akzeptabel. Das standardisierte Cronbachs  $\alpha$  fällt mit .828 (ich-Version: .830, wir-Version: .826) höher aus. Dies ist auf die im Auswertungsvorgehen (s. Abschnitt 9.2.1) aufgeführten Unterschiede in den Itemvarianzen zurückzuführen, denn das Merkmal Klassenführung weist die zweitgrößte Spannweite der Itemstandardabweichungen und somit auch der Itemvarianzen (Spannweite<sub>KF</sub> = 0.39) auf. Die mittlere Interitemkorrelation (MIC) beträgt  $MIC_{\text{insgesamt}} = MIC_{\text{ich-Version}} = MIC_{\text{wir-Version}} = .22$ . Dabei streuen die Interitemkorrelationen in einem relativ weiten Bereich von minimal  $MIC_{\text{min}} = .03$  (ich-Version:  $MIC_{\text{min}} = .02$ ; wir-Version:  $MIC_{\text{min}} = .03$ ) bis  $MIC_{\text{max}} = .70$  (ich-Version:  $MIC_{\text{max}} = .71$ ; wir-Version:  $MIC_{\text{max}} = .69$ ).

Bei der Betrachtung der Itemtrennschärfen in Abbildung 15 zeigt sich, dass fünf Items hohe Trennschärfen ( $r_{\text{itc}} > .50$ ) aufweisen – SKF7, SKF8, SKF13, SKF16 und SKF17. Diese spiegeln somit die anderen Items der Skala Klassenführung sehr gut wider. Allerdings sind darunter zwei Paare inhaltlich äquivalenter Formulierungen (Paar SKF7 – SKF13 sowie SKF8 – SKF16).

Müsste man sich aufgrund der bisherigen Befunde zwischen den beiden entscheiden, so würde die Wahl hinsichtlich des ersten Itempaares aufgrund der geringen Missinganzahl und des geringfügig höheren Trennschärfekoeffizienten auf das Item SKF7 fallen. Beim zweiten Itempaar würde das Item SKF16 aufgrund des höheren Trennschärfekoeffizienten gewählt. Beim Itempaar SKF11 – SKF4, welches im mittleren Trennschärfebereich angesiedelt ist, weist das Item SKF11 im Vergleich eine höhere Korrelation mit der Skala Klassenführung auf.

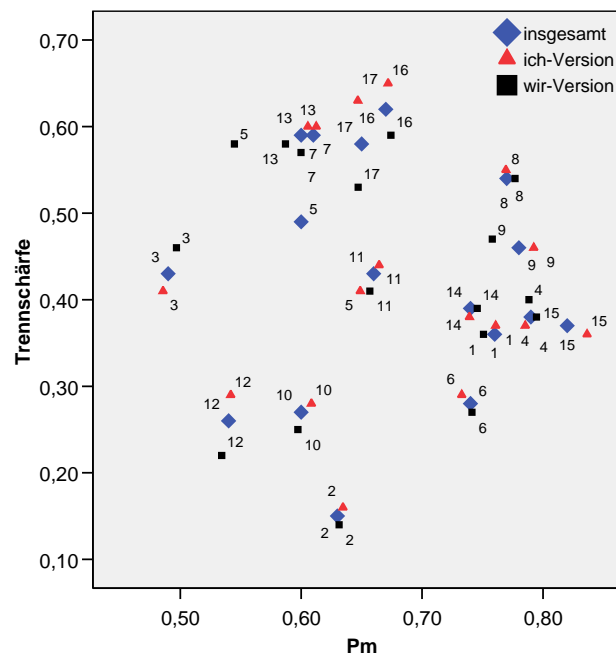


Abbildung 15: Dargestellt sind die gegen den mehrstufigen Zustimmungssindex geplotteten Trennschärfen der Klassenführungsitems in Abhängigkeit von der Fragebogenversion

Im Bereich eines geringen Trennschärfekoeffizienten liegen in steigender Reihenfolge der Trennschärfekoeffizienten SKF2, SKF12, SKF10 und SKF6. Die geringe Trennschärfe wirkt sich bei den letzten drei Items auf die Messgenauigkeit der Skala aus. So würde Cronbachs  $\alpha$  im Vergleich zum Gesamt-Cronbachs  $\alpha$  mit allen 17 Items auf .825 (ich-Version: .829, wir-Version: .822) ansteigen, wenn das Item SKF2 nicht in die Skala aufgenommen werden würde. Die Trennschärfe des Items SKF2 ist mit  $r_{itc} = .15$  (ich-Version:  $r_{itc} = .16$ , wir-Version:  $r_{itc} = .14$ ) deutlich unter .30 und hat somit insgesamt und in den beiden Fragebogenversionen den geringsten Trennschärfewert überhaupt. Die anderen drei Items liegen knapp unterhalb der .30-Grenze, wobei SKF12 in der wir-Version mit  $r_{itc} = .22$  deutlich unterhalb von .30 liegt. Werden die jeweiligen Itemschwierigkeiten bei Betrachtung der Itemtrennschärfen anhand des Selektionskennwertes mit berücksichtigt, ergeben sich dieselben Schlussfolgerungen. Auch die Analyse der Koeffizienten der ich- und wir-Version ergibt nahezu übereinstimmende Befunde und bei allen Items dieselben Schlussfolgerungen hinsichtlich der Itemauswahl. Nur das Item SKF5 variiert über die ich- und wir-Version hinweg relativ stark hinsichtlich der Trennschärfe (ich-bezogene Formulierung:  $r_{itc} = .41$ ; wir-bezogene Formulierung:  $r_{itc} = .58$ ), wobei die Streuung gleich ist und die prozentualen Kategorienhäufigkeiten nahezu gleich verteilt sind.

### 9.2.3 Skala Lernförderliches Klima

Dieser Skalenanalyse liegen in der Gesamtstichprobe 1392 (ich-Version: 698, wir-Version: 694) Schülerangaben zugrunde. Die Reliabilitätsschätzung anhand Cronbachs  $\alpha$  basierend auf 14 Items ergibt einen Koeffizienten von .860 (ich-Version: .851; wir-Version: .869), der hinsichtlich der Einstufung von Fisseni (1997) ebenfalls im mittleren Bereich liegt. Das standardisierte Cronbachs  $\alpha$  fällt mit .866 (ich-Version: .860; wir-Version: .872) etwas höher aus. Die mittlere Interitemkorrelation der Skalenitems beträgt  $MIC_{\text{insgesamt}} = .32$  ( $MIC_{\text{ich-Version}} = .30$ ;  $MIC_{\text{wir-Version}} = .33$ ). Die Präzision von Cronbachs  $\alpha$  ist mit  $P_{\alpha} = 0.01$  in der Gesamtstichprobe und den beiden Itemversionen akzeptabel.

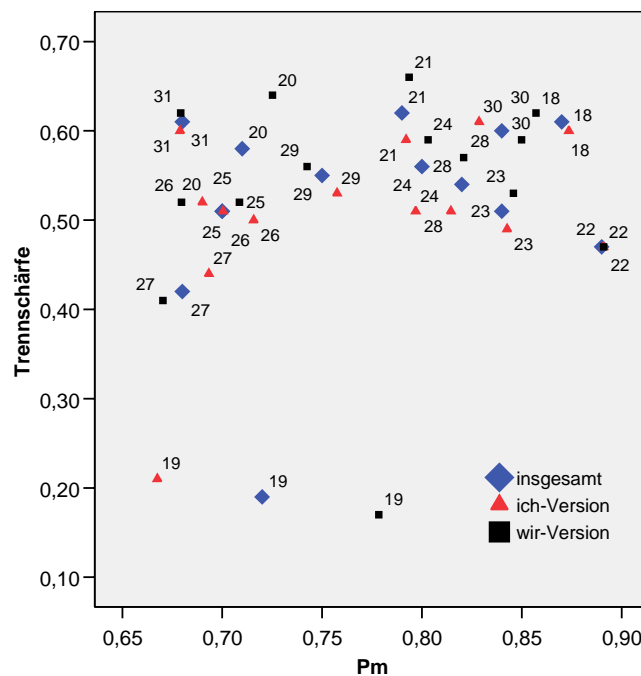


Abbildung 16: Dargestellt sind die gegen den mehrstufigen Zustimmungssindex geplotteten Trennschärfen der Items zur Erfassung des Merkmals Lernförderliches Klima in Abhängigkeit von der Fragebogenversion

Wie in Abbildung 16 ersichtlich, weist der Großteil der Items einen Trennschärfekoeffizienten von mindestens  $r_{\text{itc}} = .40$  auf. Bis auf die zwei Items SLK22 und SLK27 (und in der ich-bezogenen Formulierung noch das Item SLK23), welche einen mittleren Trennschärfekoeffizienten aufweisen, haben alle restlichen Items einen hohen, der beim Item SLK21 in der Gesamtstichprobe mit  $r_{\text{itc}} = .62$  und in der ich-Version mit  $r_{\text{itc}} = .59$  sein Maximum aufweist. In der wir-Version hat das Item SLK30 den maximalen Trennschärfekoeffizienten von  $r_{\text{itc}} = .61$ . Das Item SLK19 hat als einziges Item eine geringe Trennschärfe und einen sehr niedrigen Selektionskennwert mit  $SK_{\text{insgesamt}} = .09$  ( $SK_{\text{ich-Version}} = SK_{\text{wir-Version}} = .10$ ).



Außerdem führt die Elimination des Items sowohl in der ich- als auch der wir-Version zu einer Erhöhung von Cronbachs  $\alpha$  auf .871 in der Gesamtstichprobe. Da dieses gleichzeitig zu einer inhaltsäquivalenten Itemgruppe gehört (SLK 31 – SLK 26 – SLK19), liegt dessen Ausschluss nahe. Aufgrund des höheren Trennschärfe- und Selektionskennwertes fiel die Entscheidung bei den verbleibenden zwei Items auf SLK31.

#### 9.2.4 Skala Motivierung

1343 (ich-Version: 676, wir-Version: 667) Schülerangaben werden bei dieser Analyse verwendet. Cronbachs  $\alpha$  beträgt für die 13 Items .861 (ich-Version: .858; wir-Version: .865) und ist somit ein mittlerer Koeffizient (Fisseni, 1997). Etwas höher ist das standardisierte Cronbachs  $\alpha$  mit .862 (ich-Version: .858; wir-Version: .866). Die mittlere Interitemkorrelation beträgt  $MIC_{\text{insgesamt}} = MIC_{\text{ich-Version}} = .32$ ,  $MIC_{\text{wir-Version}} = .33$  und die Präzision von Cronbachs  $\alpha$  ist mit  $P_{\alpha} = 0.01$  in der Gesamtstichprobe und den beiden Itemversionen akzeptabel.

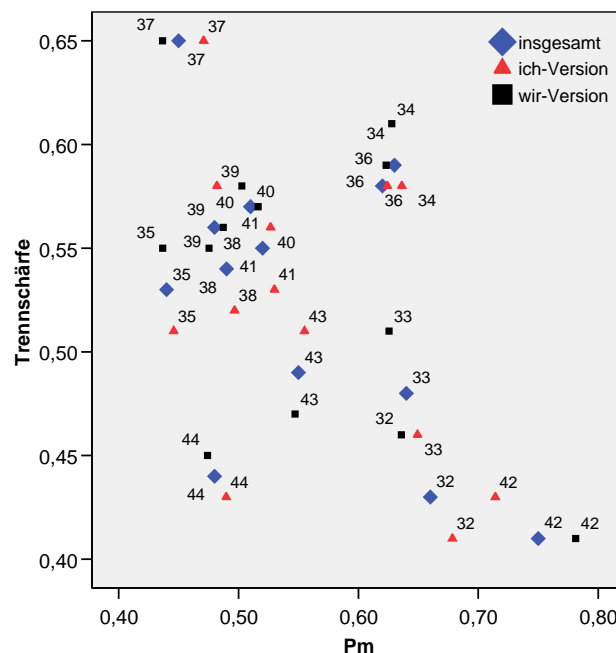


Abbildung 17: Dargestellt sind die gegen den mehrstufigen Zustimmungssindex geplotteten Trennschärfen der Motivierungsitems in Abhängigkeit von der Fragebogenversion

Hinsichtlich der Trennschärfe ist in Abbildung 17 auffällig, dass alle Items einen Koeffizienten von mindestens  $r_{itc} = .40$  haben – mit einer hohen Einstufung ( $r_{itc} > .50$ ) bei acht und einer mittleren ( $.40 < r_{itc} < .50$ ) bei fünf Items bzw. in der ich- und wir-Version jeweils bei sechs Items. Der Trennschärfe- und Selektionskennwert ist beim Item SM37 mit  $r_{itc} = .65$  und  $SK = .33$  am höchsten. Die übrigen Items liegen mit der Ausnahme des Items

SM32 in der ich-Version ( $SK = .19$ ) hinsichtlich des Selektionskennwertes in einem mittleren Bereich zwischen  $.20 < SK < .30$ . Somit sind alle Items relativ prototypisch für die Motivierungsskala. Die Elimination nur eines Items (SM35) wird aus dem Grund inhaltlicher Ähnlichkeit und eines höheren Trennschärfe- und Homogenitätswertes im Vergleich zum Item SM39 nahe gelegt.

### 9.2.5 Skala Klarheit und Verständlichkeit

Die Analyse basiert in der Gesamtstichprobe auf 1390 (ich-Version: 707, wir-Version: 683) Schülerangaben. Die 16 Items weisen einen mittelhohen Reliabilitätskoeffizienten von Cronbachs  $\alpha = .848$  (ich-Version:  $.840$ ; wir-Version:  $.858$ ) auf. Mit  $.851$  (ich-Version:  $.842$ ; wir-Version:  $.863$ ) ist das standardisierte Cronbachs  $\alpha$  etwas höher. Die mittlere Interitemkorrelation beträgt  $MIC_{\text{insgesamt}} = .26$  ( $MIC_{\text{ich-Version}} = .25$ ,  $MIC_{\text{wir-Version}} = .28$ ) und die Präzision von Cronbachs  $\alpha$  ist mit  $P_{\alpha} = .008$  sehr gut, sowohl in der Gesamtstichprobe als auch in den beiden Itemversionen. Dies ist allerdings nur eine notwendige und keine hinreichende Bedingung für die Eindimensionalität der Skala.

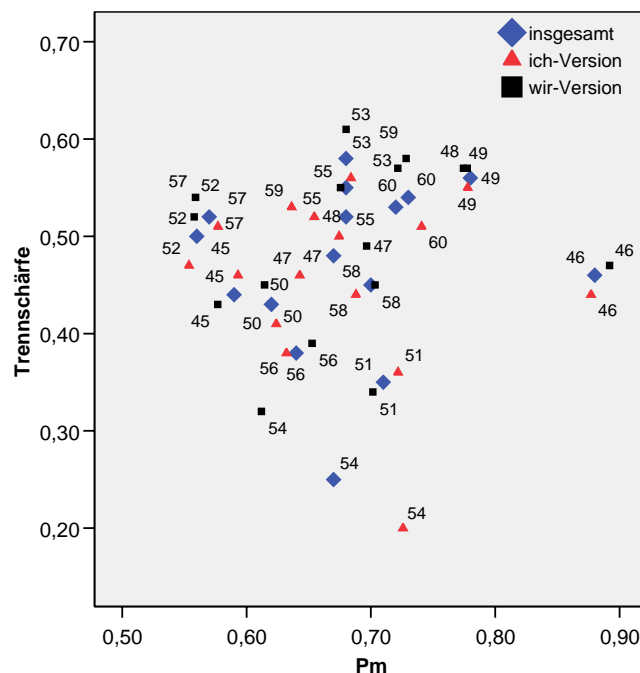


Abbildung 18: Dargestellt sind die separat für die Fragebogenversionen sowie für die Gesamtstichprobe geplotteten mehrstufigen Zustimmungsindizes und die Trennschärfen der Items zur Erfassung der Klarheit und Verständlichkeit

Ein Blick auf die Trennschärfen in Abbildung 18 zeigt, dass sieben Items eine hohe Trennschärfe ( $r_{\text{itc}} > .50$ ) haben. Das Item SKV54 hat in der Gesamtstichprobe und der ich-Version eine geringe Trennschärfe von  $r_{\text{itc}} < .30$  (wir-Version:  $r_{\text{itc}}(\text{SKV54}) = .32$ ). Ferner

erhöht sich Cronbachs  $\alpha$  geringfügig in der ich-Version und der Gesamtstichprobe beim Herausnehmen dieses Items. In Übereinstimmung mit diesen Befunden rät auch der Selektionskennwert zu dessen Ausschluss und zusätzlich noch zur Elimination von Item SKV51 und SKV56 (und in der ich- bzw. wir-Version auch noch von Item SKV45 und SKV50). Das Item SKV45 ist bezüglich des Inhalts äquivalent zu SKV52 und würde im direkten Vergleich aufgrund der geringeren Kennwertkoeffizienten so abschneiden, dass SKV52 im Fragebogen verbleiben sollte.

### 9.2.6 Skala Aktivierung

Insgesamt fußt diese Analyse der Aktivierungsskala auf 1355 (ich-Version: 690; wir-Version: 665) Schülerangaben. Die Reliabilitätsschätzung der 18 Aktivierungsitems ergibt einen Koeffizienten von Cronbachs  $\alpha = .851$  (ich-Version: .852, wir-Version: .854) und liegt gemäß Fisseni (1997) im mittleren Bereich. Das standardisierte Cronbachs  $\alpha$  fällt mit .851 (ich-Version: .852, wir-Version: .853) etwas höher aus. Die mittlere Interitemkorrelation beträgt  $MIC_{\text{insgesamt}} = MIC_{\text{ich-Version}} = MIC_{\text{wir-Version}} = .24$  und die Präzision von Cronbachs  $\alpha$  ist mit  $P_{\alpha} = .007$  im Vergleich zu den anderen vier Skalen am geringsten.

Die Trennschärfekoeffizienten sind wie in Abbildung 19 bei fünf Items hoch ( $r_{\text{itc}} > .50$ ). Die Unterschiede hinsichtlich des Trennschärfewertes zwischen der ich- und wir-Version sind bei den Aktivierungsitems am stärksten. So ist die Trennschärfe des Items SA78 in der ich-Version und der Gesamtstichprobe mit  $r_{\text{itc}} = .28$  gering und Cronbachs  $\alpha$  würde bei einem Itemausschluss geringfügig erhöht, wohingegen es in der wir-Version einen deutlich höheren und hinsichtlich der Einstufung mittleren Trennschärfekoeffizienten von .41 hat und bei Elimination zur Reduktion des Skalenreliabilitätswerts führt. Der Selektionskennwert legt allerdings in beiden Fällen den Ausschluss nahe. Auch das Item SA68 differiert hinsichtlich des Trennschärfekoeffizienten in der Einstufung in Abhängigkeit von der Fragebogenversion. In der wir-Version spiegelt es die anderen Skalenitems mit einer Trennschärfe von  $r_{\text{itc}} = .24$  nur geringfügig wider und drückt dadurch die Skalenreliabilität, wohingegen dieses in der ich-Version mit  $r_{\text{itc}} = .37$  bzw. in der Gesamtstichprobe mit  $r_{\text{itc}} \text{ SA68} = .31$  eine mittlere Passungsgüte aufweist und in beiden Fällen einen positiven Beitrag zur Skalenreliabilität hat. Der Selektionskennwert würde auch zum Ausschluss der Items SA64, SA67, SA68 und SA75 führen sowie zusätzlich in der wir-Version zur Elimination des Items SA69 und der ich-Version von SA62.

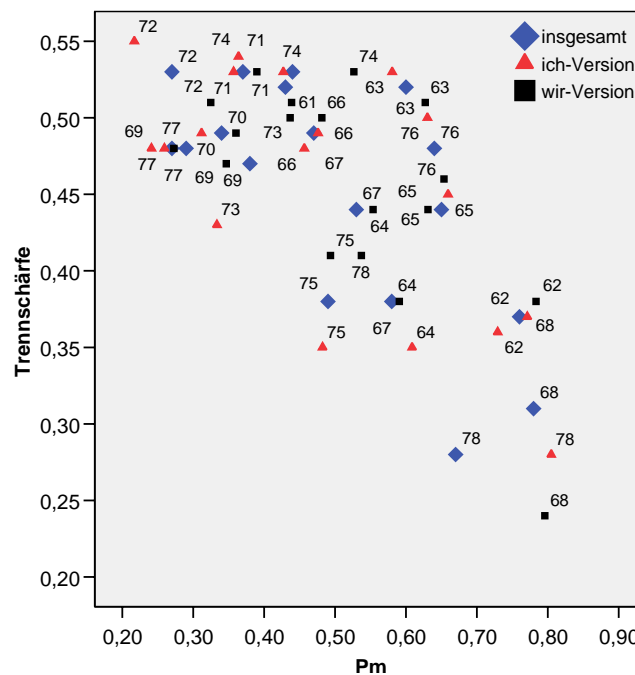


Abbildung 19: Plot der mehrstufigen Zustimmungsindizes der Aktivierungsitems gegen die Trennschärfewerte für die beiden Fragebogenversionen sowie in der Gesamtstichprobe

### 9.2.7 Zusammenfassende Bewertung

Insgesamt zeigen sich durchweg positive Trennschärfekoeffizienten, die bei 36 Items in der Gesamtstichprobe (ich-Version: 37 Items; wir-Version: 41 Items) gemäß Fisseni (1997) mit einem Wert von  $r_{itc} > .50$  als hoch klassifiziert werden können. Die Trennschärfewerte nehmen dabei mit zunehmenden Zustimmungsindizes entsprechend der testtheoretischen Annahmen bei den Unterrichtsmerkmalen Klassenführung, Motivierung und Aktivierung ab. Dieser Sachverhalt zeigt sich bei den Aktivierungsitems am deutlichsten.

Die Betrachtung der Cronbachs Alpha-Werte in der Übersicht der Tabelle 14 zeigt, dass alle Skalen gemäß der Klassifikation von Fisseni (1997) einen mittleren Wert zwischen  $.80 \leq r_{it} \leq .90$  aufweisen. Die Skala Motivierung hat die höchsten Reliabilitätskoeffizienten in der Gesamtstichprobe, gefolgt von der Skala Lernförderliches Klima. Mit etwas Abstand folgen die beiden Skalen Klarheit und Verständlichkeit sowie Aktivierung und mit deutlichem Abstand Klassenführung. Die Interitemkorrelationskoeffizienten der Klassenführungsitems streuen am stärksten, so beträgt das Bereichsmaß zwischen  $MIC_{MAX} - MIC_{MIN} = .67$ . Eine ähnlich hohe Streuung der Interitemkorrelationen zeigen die Items zur Erfassung des Lernförderlichen Klimas. Danach folgen die Bereichsmaße  $MIC_{MAX} - MIC_{MIN}$  (Aktivierung) = .43 und  $MIC_{MAX} - MIC_{MIN}$  (Klarheit und Verständlichkeit) = .48 und schließlich  $MIC_{MAX} - MIC_{MIN}$  (Motivierung) = .39. Die Skala Klassenführung hat

somit die höchste Streuung der Interitemkorrelationskoeffizienten und den geringsten Reliabilitätswert. Dies wird in Teilen durch die Unterschiede in den Itemvarianzen verursacht. Denn wie im Abschnitt 9.1.5 gezeigt, besitzt das Merkmal Klassenführung mit 0.39 die zweitgrößte Spannweite<sub>KF</sub> der Itemstandardabweichungen. Dies zeigt sich auch darin, dass der Unterschied zwischen dem standardisierten und unstandardisierten Cronbachs  $\alpha$  mit  $> .10$  Abweichungseinheiten am größten ist. Beim Lernförderlichen Klima fällt der  $P_\alpha = .012$  am höchsten aus, dies ist ein Indiz dafür, dass die Skala nicht eindimensional ist.

Tabelle 14: *Reliabilitäts- und Homogenitätswerte der fünf Skalen in Abhängigkeit von der Fragebogenversion*

Kennwert	Version	Skala				
		Klassen- führung	Lernförder- liches Klima	Moti- vierung	Klarheit und Ver- ständlichkeit	Aktivierung
<b>Itemanzahl</b>	alle	17	14	13	16	18
<b>Cronbachs <math>\alpha</math></b>	<b>insg.</b>	<b>.814</b>	<b>.860</b>	<b>.861</b>	<b>.848</b>	<b>.851</b>
	ich	.818	.851	.858	.840	.852
	wir	.811	.869	.865	.858	.854
<b>standard. Cronbachs <math>\alpha</math></b>	<b>insg.</b>	<b>.828</b>	<b>.866</b>	<b>.862</b>	<b>.851</b>	<b>.851</b>
	ich	.830	.860	.858	.842	.852
	wir	.826	.872	.866	.863	.853
<b>Präzision von Alpha</b>	<b>insg.</b>	<b>.010</b>	<b>.012</b>	<b>.010</b>	<b>.008</b>	<b>.007</b>
	ich	.010	.011	.010	.008	.007
	wir	.011	.013	.010	.008	.007
<b>Mittlere Interitem- korrelation</b>	<b>insg.</b>	<b>.22</b>	<b>.32</b>	<b>.32</b>	<b>.26</b>	<b>.24</b>
	ich	.22	.30	.32	.25	.24
	wir	.22	.33	.33	.28	.24
<b>Mittlerer Itemmit- telwert</b>	<b>insg.</b>	<b>3.01</b>	<b>3.33</b>	<b>2.67</b>	<b>3.04</b>	<b>2.49</b>
	ich	3.02	3.32	2.67	3.03	2.45
	wir	2.99	3.33	2.66	3.05	2.53

Die aufgeführten Befunde zeigen sich im Wesentlichen auch in der ich- und wir-Version, wobei die Skala Lernförderliches Klima hinsichtlich der Rangreihe der Cronbachs Alpha-Werte abweicht. In der wir-Version weist es den höchsten Reliabilitätswert auf, in der ich-Version den dritthöchsten.

Hinsichtlich möglicher Schlussfolgerungen in Bezug auf die Selektion auf Itemebene zeigt sich, dass in jedem Fall inhaltsäquivalente Items ausgeschlossen werden sollten, um die Skalenreliabilität nicht künstlich zu erhöhen und inhaltliche Redundanz zu vermeiden. Dazu werden der Trennschärfe-, der Selektionskennwert und die Homogenitätsindizes herangezogen, die für den Ausschluss der Items SKF4, SKF8, SKF13, SLK19, SLK26, SM35 und SKV45 sprechen. Die Analyseergebnisse regen ferner den Ausschluss der Items

SKF2, SKF6, SKF10, SKF12, SKV54 und SA78 an. Im Folgenden soll überprüft werden, inwiefern sich diese Befunde auch faktorenanalytisch zeigen.

### **9.3 Faktorielle Validitätsanalysen**

Nach einem Überblick zum Auswertungsvorgehen (Abschnitt 9.3.1) erfolgt die faktorenanalytische Datenbetrachtung in zwei Schritten: (1) Exploratorisch, um die Faktorenstruktur datengeleitet zu entwickeln (Abschnitt 9.3.2) und (2) konfirmatorisch zur Überprüfung der Itempassung in Bezug auf die a priori postulierten Skalen (Abschnitt 9.3.3). Im Abschnitt 9.3.4 werden wesentliche Ergebnisse zusammengeführt.

#### **9.3.1 Auswertungsvorgehen**

Im ersten Schritt werden die Daten der Gesamtstichprobe und separat für die ich- und wir-Version einer exploratorischen Faktorenanalyse unterzogen, um anhand korrelativer Itemähnlichkeiten die vorliegende Faktorenstruktur und die Itempassung bestimmen zu können. Aufgrund der Tatsache, dass hoch korrelierte Faktoren bzw. Faktoren höherer Ordnung angenommen werden, wird die oblique Rotationstechnik gewählt, die gleichzeitig das Erreichen einer Einfachstruktur erleichtert. Aufgrund des Fokus auf eine Datenreduktion erfolgt eine Hauptkomponentenanalyse (Bühner, 2006). Die Autorin ist sich der Abhängigkeit der Analyseergebnisse von der Methode jedoch bewusst, daher kommen zusätzlich Sensitivitätsanalysen in Bezug auf verschiedene Rotationstechniken und Faktorenanalysemethoden zum Einsatz.

Im zweiten Schritt erfolgt die Durchführung konfirmatorischer Faktorenanalysen, bei denen im Gegensatz zur exploratorischen Faktorenanalyse Itemladungen nur auf dem postulierten Faktor zugelassen werden. Dadurch können Aussagen zur Eindimensionalität der Faktoren getroffen und Fehlerkorrelationen zwischen inhaltsäquivalenten Items spezifiziert werden. Die Passungsgüte bzw. der Modellfit hinsichtlich der theoretisch postulierten Modellstruktur zur empirisch beobachteten Datenstruktur wird anhand von Gütekriterien beurteilt, die das Mess- bzw. Strukturmodell insgesamt betrachten (Backhaus et al., 2006). Gemäß Beauducel und Wittmann (2005) können dabei die folgenden Fit-Indizes herangezogen werden:  $\chi^2$ -Wert und der dazugehörige p-Wert, der Comparative-Fit-Index (CFI), der RMSEA (Root-Mean-Square-Error of Approximation), welcher inferenzstatistische Betrachtungen erlaubt und der in Kombination mit dem RMSEA empfohlene Standardized-Root-Mean-Residual (SRMR). Diese Indizes werden im Anhang D.3.3 hinsichtlich ihrer Funk-

tion und Cut-off-Werte beschrieben. Für den vorliegenden Datensatz erweist sich der  $\chi^2$ -Wert als ungeeignete Teststatistik, da dieser sehr sensitiv gegenüber Veränderungen des Stichprobenumfangs und Abweichungen von der Normalverteilungsannahme ist (Backhaus et al., 2006; Schermelleh-Engel, Moosbrugger & Müller, 2003). Bei großen Stichproben ( $n > 200$ ) wird dieser Wert in der Regel aufgrund der Teststärke der  $\chi^2$ -verteilten Prüfgröße immer signifikant. Deshalb werden im Wesentlichen der CFI, der RMSEA und SRMR herangezogen. Bezüglich der Schätzmethode wird entsprechend den Ausführungen im Abschnitt 8.2 das ML-Schätzverfahren unter Verwendung einer Bollen-Stine-Bootstrap-Korrektur zur Identifikation von Über- vs. Unterschätzungen verwendet.

### 9.3.2 Exploratorische Faktorenanalyse

Werden die Schülerangaben einer exploratorischen Faktorenanalyse (HKA, Oblimin-Rotation) unterzogen, ergeben sich die im Folgenden dargestellten *Faktorenextraktionsmöglichkeiten*. Das Kaiser-Kriterium, welches die Extraktion von Faktoren erlaubt, die mindestens genauso viel Gesamtvarianz aufklären, wie ein Ausgangsitem beisteuert und somit einen Eigenwert größer als 1.0 haben, legt die Extraktion von 16 Faktoren nahe.

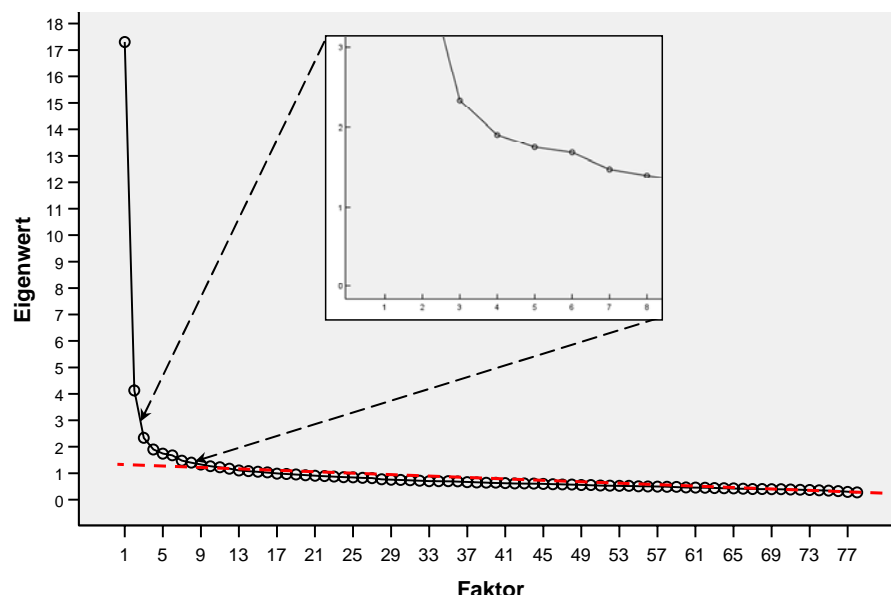


Abbildung 20: Scree-Plot der 78 Faktoren gegen die Eigenwerte, wobei die Hilfslinie eine Gerade darstellt, die durch die letzten extrahierten Faktoren gelegt wird, die von Cattell (1966) als ‚Geröll‘ bezeichnet werden

Zieht man das Scree-Kriterium zur Entscheidung hinsichtlich der Faktorenextraktion hinzu, so würden alle folgenden Faktoren ab dem Punkt extrahiert, an dem die Verbindungslinie der Eigenwerte der Faktoren in dem Scree-Plot (s. Abbildung 20) deutlich über die Gerade

steigt, die durch die Eigenwerte der letzten extrahierten Faktoren gelegt wird. Die Interpretation des ‚deutlichen Steigens‘ lässt einen Spielraum von 1, 2, 3 oder sogar 6 extrahierten Faktoren zu, wenn man den Sprung im Anstieg vom 7. auf den 6. Faktor vor dem Hintergrund des nahezu konstanten Anstiegs der Eigenwerte ab dem 78. Faktor mit einbezieht.

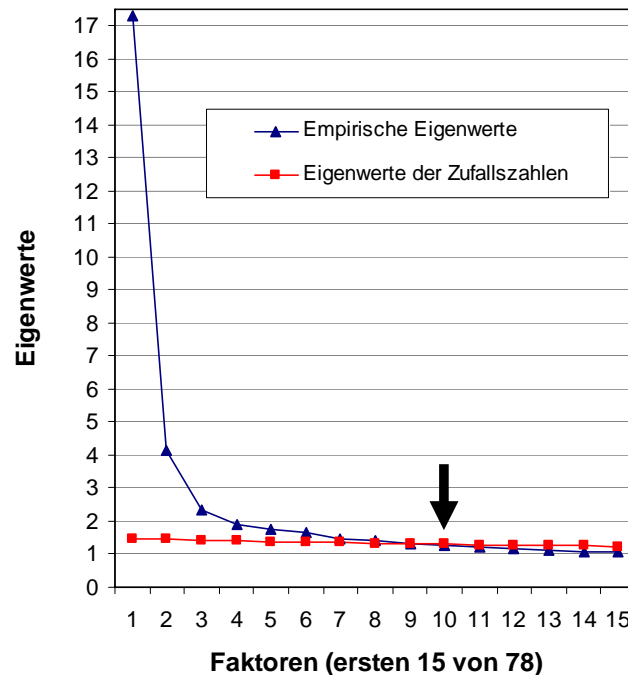


Abbildung 21: Vergleich der empirischen und des aufgrund der Parallelanalyse nach Horn (1965) ermittelten Eigenwertverlaufs der Zufallszahlen, wobei der Pfeil den 10. Faktor markiert, bei dem der empirische Eigenwert zum ersten Mal kleiner ist als der auf der Grundlage von Zufallszahlen ermittelte Eigenwert

Bei der Parallelanalyse erfolgt die Faktorenextraktion nur dann, wenn der empirisch beobachtete Eigenwert größer ist als der Eigenwert, der anhand zufälliger Daten erzeugt wird (Bortz, 1999). Die Stichprobe für die Ermittlung des zufälligen Eigenwertverlaufs entspricht dabei der empirischen Stichprobe hinsichtlich der Item- und Fallanzahl (O'Connor, 2000). Der zufällige Eigenwertverlauf wurde mit Hilfe der Software von Watkins (2004) mit einer Spezifikation von 78 Variablen, 1525 Fällen und 100 Replikationen erstellt (s. CD-Rom). Die Parallelanalyse legt die Extraktion von 9 Faktoren nahe, da ab dem 10. Faktor der empirisch beobachtete Eigenwert erstmals größer ist als der aufgrund der Zufallszahlen ermittelte Eigenwert. Dies ist in Abbildung 21 dargestellt. Allerdings ist bei der Parallelanalyse keiner der Faktoren auf einem Niveau von  $p < .05$  inferenzstatistisch bedeutsam.



Zusammenfassend lässt sich festhalten, dass die verschiedenen Extraktionskriterien unterschiedliche Resultate ergeben: Das Kaiser-Kriterium legt 16 Faktoren, der Scree-Test 1, 2, 3 oder 6 Faktoren, die Parallelanalyse 9 Faktoren nahe und gemäß theoretischer Vorgaben müssten 5 Faktoren extrahiert werden.

Bei der Entscheidung, welche Anzahl an Faktoren extrahiert werden soll, gilt es neben der inhaltlichen Plausibilität der Faktoren zu beachten, dass Zwick und Velicer (1986) zufolge das Kaiser-Kriterium die tatsächliche Faktorenanzahl tendenziell überschätzt. Die Parallelanalyse sollte dem Scree-Kriterium vorgezogen werden, da dieses subjektiv ist. Hinsichtlich des objektiveren Parallelanalyse-Kriteriums hat allerdings Beauducel (2001) in einer Simulationsstudie gezeigt, dass die Faktorenanzahl bei einer starken ersten Hauptkomponente überschätzt wird. Dies trifft bei dem vorliegenden Datensatz zu, da die erste Hauptkomponente vor der Rotation mit 17.3 einen sehr hohen Eigenwert hat und somit 23% der gesamten Itemvarianz aufklärt. Außerdem führt eine große Stichprobenanzahl bei der Parallelanalyse dazu, dass der auf den Zufallszahlen beruhende Eigenwertverlauf flacher wird und dementsprechend die Extraktion von mehr Faktoren nahe gelegt wird, als es bei dem gleichen empirischen Eigenwertverlauf bei einer kleineren Stichprobengröße der Fall wäre. Die Extraktion von 9 Faktoren würde zu Skalen mit nur zwei Items führen, was hinsichtlich des Reliabilitätsaspekts zu wenig wäre. Aufgrund der starken Heterogenität in Bezug auf die Faktorenextraktion werden mehrere Faktorenlösungen hinsichtlich ihrer Faktoren in der Gesamtstichprobe interpretiert. Dazu werden Markiervariablen herangezogen, d. h. Items mit hohen Faktorladungen. Die Ergebnisse sind im Anhang D.3.1 näher dargestellt und zeigen, dass sich über die verschiedenen Faktorenlösungen hinweg nur das Merkmal Klarheit und Verständlichkeit nicht gut in den Daten replizieren lässt, wobei sich die Anzahl eindeutig replizierbarer Faktoren in Abhängigkeit von der Faktorenlösung unterscheidet. In der 4- und 5-Faktorenlösung laden die Items zum Erfassen der Klarheit und Verständlichkeit im Wesentlichen auf dem Motivierungsfaktor, wohingegen sie in der 6-Faktorenlösung einen eigenständigen Faktor bilden bzw. in den beiden Fragebogenversionen einen gemeinsamen Faktor mit den Aktivierungsitems. In Anbetracht dieser Ergebnisse wird die Extraktion von 5 Faktoren als guter Kompromiss zwischen einer akzeptablen Faktorsättigung und der Modellsparsamkeit gewählt. Die 5-Faktorenlösung weist zudem einen pädagogisch-didaktischen Vorteil durch die Übereinstimmung mit den im Orientierungsrahmen Schulqualität in Rheinland-Pfalz veröffentlichten Faktoren auf (MBWJK, 2007).

Die ausführlichen Ergebnisse der 5-Faktorenlösung für die Gesamtstichprobe und für die ich- und wir-Version finden sich im Anhang D.3.2. Es zeigt sich ein starker Generalfaktor – die erste Hauptkomponente klärt vor der Rotation 23% der Itemvarianz in der Gesamtstichprobe auf. Diese Hauptkomponente könnte eine kollektive Unterrichtswahrnehmung in Bezug auf die Qualität der Lehr-Lernprozesse abbilden. Das Faktorladungsmuster zeigt, dass die Faktoren Klassenführung, Aktivierung, Lernförderliches Klima und Motivierung eine relativ gute Einfachstruktur aufweisen, lediglich der Faktor Klarheit und Verständlichkeit zerfällt stark. In der wir-Version zeigt sich die 5-Faktorenstruktur deutlicher als in der ich-Version. Nur die sieben Items SKF2, SKF6, SKF10, SKF12, SKF15, SA76 und SA78 (s. Anhang D.3.2) weisen keine substanziellen Ladungen größer .30 auf. 38 Items laden bedeutsam auf dem jeweiligen a priori postulierten Faktor und sind hinsichtlich ihrer Anzahl annähernd gleich auf die vier Faktoren Klassenführung, Lernförderliches Klima, Motivierung und Aktivierung verteilt. Die Faktoreninterkorrelation beträgt in der Gesamtstichprobe zwischen  $r_{\text{Faktor A, Faktor M}} = -.44$  und  $r_{\text{Faktor KF, Faktor LK}} = .36$ , in der ich-Version zwischen  $r_{\text{Faktor LK, Faktor M, KV}} = -.39$  und  $r_{\text{Faktor KF, Faktor M, KV}} = .37$  und in der wir-Version zwischen  $r_{\text{Faktor M, Faktor A}} = -.40$  und  $r_{\text{Faktor KF, Faktor LK}} = .36$ . In einer Sensitivitätsanalyse erweisen sich die faktorenanalytischen Ergebnisse über verschiedene Rotationstechniken (Oblimin, Varimax, Promax) und Faktorenanalysemethoden (insbesondere HKA, HAA, z.T. auch ML) als recht stabil.

Die exploratorischen Faktorenanalysen legen somit die Elimination einer Reihe von Klassenführungsitems (2, 6, 10, 12 und 15) sowie die der Items SA76 und SA78 nahe. Diese Befunde stehen bis auf die Items SKF15 („Die Klassenregeln waren mir/uns bekannt“) und SA76 („Der Lehrer hat Fragen gestellt, die mich zum Nachdenken angeregt haben“) im Einklang mit den Ergebnissen der Item- und Reliabilitätsanalysen.

### 9.3.3 Konfirmatorische Faktorenanalyse

Bei der Durchführung der konfirmatorischen Faktorenanalyse wird im ersten Schritt das Gesamtmodell spezifiziert, welches alle 78 Items zur Erfassung der Unterrichtswahrnehmung als manifeste Variablen und die jeweiligen Unterrichtsmerkmale als latente Konstrukte umfasst (s. Abbildung 22). Dabei werden zusätzlich Fehlerkorrelationen zwischen den inhaltsäquivalenten Variablen und aufgrund der Annahme, dass die Faktoren nicht unabhängig voneinander sind, Interkorrelationen zwischen den fünf Konstrukten zugelassen. Zunächst soll die Güte des Modells anhand der im Abschnitt 9.3.1 beschrie-

benen Fit-Indizes und der im Anhang D.3.3 aufgeführten Cut-off-Werte beurteilt werden. Der CFI-Wert ist mit .803 nicht akzeptabel. Allerdings erhalten generell sparsame Modelle, die mit wenigen Parametern auskommen, einen höheren CFI-Wert; diese Bedingung ist im vorliegenden Kontext der Überprüfung eines umfangreichen Itempools nicht gegeben ist. Der RMSEA weist als inferenzstatistischer Fit-Index mit .041 einen guten Fit auf. Das Konfidenzintervall CI des RMSEA schließt zwar mit  $CI = [.040; .042]$  den Wert 0 nicht ein, dennoch liegt der RMSEA deutlich unterhalb der Cut-off-Grenze von .05. Auch der SRMR erfüllt die Anforderungen eines guten Fits. Der Vollständigkeit halber wird an dieser Stelle auch der signifikante  $\chi^2$ -Wert von 10454.80\*\*\* ( $df = 2906$ ) berichtet, der aufgrund der großen Stichprobe jedoch nicht hinsichtlich der Beurteilung des Modellfits hinzugezogen wird. Die Anwendung der Bollen-Stine-Bootstrap-Korrektur erhöht den p-Wert des  $\chi^2$ -Wertes auf  $p = .005$ . Die Verletzung der Verteilungsannahme wirkt sich somit zwar auf den p-Wert aus, ist aber in Einklang mit der relativen Robustheit des ML-Schätzers nach Bollen (1989) nicht sehr stark. Die Anpassungsgüte der theoretischen Modellstruktur an die empirischen Daten kann hinsichtlich des inferenzstatistischen Maßes RMSEA und des SRMR somit als gut beurteilt werden.

Die Inspektion der Parameterschätzungen der standardisierten Regressionsgewichte soll Aufschluss über die dimensionsanalytische Itemeignung geben. Diese sind ebenfalls im Modell in Abbildung 22 dargestellt. Alle Schätzungen sind kleiner 1.0. Dies ist positiv, da der gegenteilige Fall ein Hinweis auf Schätzprobleme sein könnte. Die standardisierten Regressionsgewichte aller beobachteten Variablen bzw. Items sind auf einem  $p < .001$ -Niveau signifikant. Allerdings sind diese Signifikanzwerte aufgrund der Stichprobengröße mit Vorsicht zu interpretieren (Byrne, 2001).

Jedes Item lädt auf nur einem Faktor, deshalb entsprechen die geschätzten partiellen standardisierten Regressionsgewichte den Korrelationen zwischen Item und Faktor. Das Quadrat dieser Korrelationen gibt den durch die Faktoren aufgeklärten Varianzanteil der Items an und entspricht der Itemkommunalität (Bühner, 2006). Die standardisierten Regressionsgewichte nehmen Werte von .149 (SLK19) bis .708 (SM37) an. Somit beträgt der aufgeklärte Itemvarianzanteil mindestens 2% und maximal 50%. Die Items weisen mehrheitlich Regressionsgewichte von .40 bis .60 auf. Die standardisierten Fehlervarianzanteile (entspricht  $1-R^2$ ) sind bei allen 78 Items auf einem  $p < .001$ -Niveau signifikant. Somit könnte ein erheblicher Teil dieser Fehlervarianz entweder unsystematisch sein, durch

andere Faktoren z. B. auf einer höheren Ebene aufgeklärt werden oder itemspezifisch sein (Byrne, 2001).

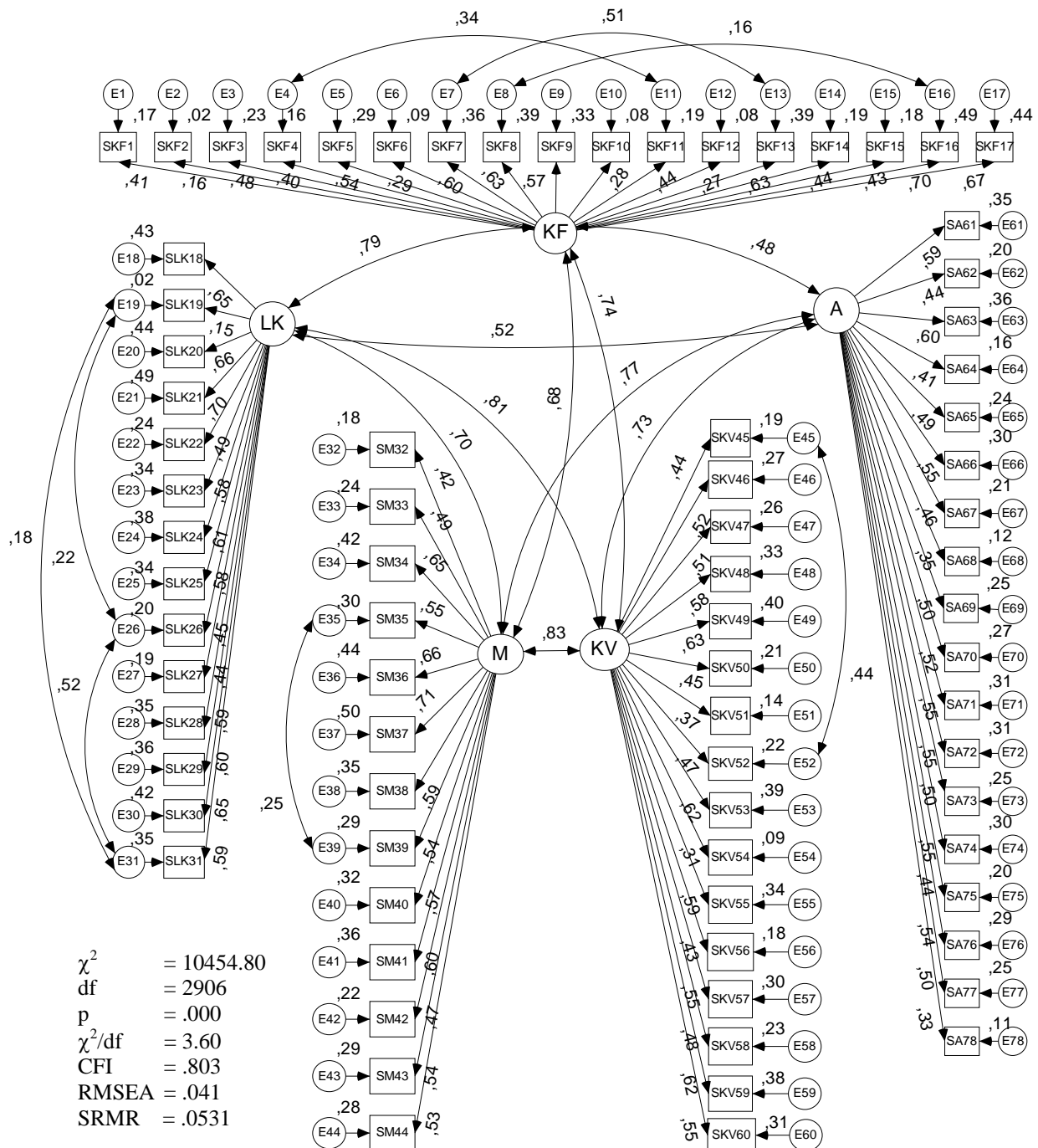


Abbildung 22: Dargestellt ist das a priori postulierte 5-Faktoren-Mess- bzw. Strukturmodell der 78 Items zur Unterrichtswahrnehmung mit den entsprechenden Fit-Indizes

Der Blick auf Maßnahmen zur Verbesserung des Modellfits anhand von Modellmodifikationen zeigt, dass v. a. Messfehler-Interkorrelationen mit anderen Faktoren nahegelegt werden. Der  $\chi^2$ -Wert würde sich relativ stark, jedoch nicht entscheidungsrelevant ver-

ringern, wenn die inhaltlich plausiblen Fehlerkorrelationen, z. B. der Items SA61 („Lehrkraft hat Hinweise gegeben, wie ich besser lernen kann“) und SA66 („Lehrkraft hat Hinweise zur Zeitplanung und Zeiteinteilung beim Lernen gegeben“) oder eine skalenübergreifende Korrelation zwischen SKF5 („Ich war die ganze Stunde über aufmerksam und konzentriert“) und SM44 („Ich war während des Unterrichts aktiv bei der Sache“) zugelassen würden. Dies ist ein Hinweis auf einen relativ großen systematischen Fehlervarianteanteil, der sich auch in einem starken Generalfaktor der Unterrichtswahrnehmung niederschlagen könnte. Diese Überlegung wird am Ende des Abschnitts erneut aufgegriffen.

In Tabelle 15 sind die Items mit einem aufgeklärten Varianzanteil von kleiner 15% aufgeführt. Die Kommunalitäten können als Mindestschätzungen der Reliabilitäten der Items erachtet werden und fallen bei diesen Items entsprechend gering aus. Dies liegt u. a. daran, dass die Reliabilität eines Items zwangsläufig nicht so hoch sein kann, wie die einer Skala. Die niedrigen Kommunalitäten könnten auch ein Hinweis auf vorhandene Heterogenität der Items sein. Aufgrund dieses Befundmusters wird der Ausschluss zumindest der Items mit einer Varianzaufklärung kleiner 10% nahe gelegt.

Tabelle 15: *Items mit standardisierten Regressionsgewichten  $\leq .35$*

Item	stand. Regressionsgewicht	Kommunalität	$R^2_{\text{Item, Faktor}}$ in %	$1-R^2_{\text{Item, Faktor}}$ in %
SLK19	.149	.022	2,2	97,8
SKF2	.157	.025	2,5	97,5
SKF12	.274	.075	7,5	92,5
SKF10	.283	.080	8,0	92,0
SKF6	.293	.086	8,6	91,4
SKV54	.307	.094	9,4	90,6
SA78	.328	.108	10,8	89,2
SA68	.350	.123	12,3	87,7

Abschließend soll die Faktorenebene als Zusatzinformation in Bezug auf die Fragebogengüte betrachtet werden. Die Spezifikation der Faktoren in Einfaktor-Modellen zeigt die in Tabelle 16 angegebenen Fit-Indizes. Demnach weisen die Faktoren Klarheit und Verständlichkeit, Aktivierung sowie Klassenführung in Bezug auf den RMSEA und SRMR akzeptable Fits auf. Der CFI-Wert ist entsprechend den Ergebnissen auf Gesamtmodellebene nicht akzeptabel. Dieses Befundmuster entspricht den Ergebnissen der exploratischen Faktorenanalyse hinsichtlich der Varianzaufklärung der einzelnen Faktoren (s. Abschnitt 9.3.2) und den sehr guten Präzisionswerten von Cronbachs  $\alpha$  als Indikatoren für eine Eindimensionalität bei dem Merkmal Aktivierung und Klarheit und Verständlichkeit und dem

akzeptablen Präzisionswert bei der Klassenführung (s. Abschnitt 9.3.2). Lernförderliches Klima und Motivierung weisen lediglich hinsichtlich des SRMR einen akzeptablen Wert auf – entsprechend einem geringen Varianzaufklärungsanteil in den exploratorischen Analysen und einem geringen Präzisionswert von Cronbachs  $\alpha$ .

Tabelle 16: *Modell-Fit-Indizes zu verschiedenen Modellvarianten*

Modellvariante		Modell-Fit-Indizes der konfirmatorischen FA				
		$\chi^2$	df	CFI	RMSEA	SRMR
Einfaktor-Modelle	Klarheit & Verständlichkeit	894.44	104	.858	<b>.071</b>	<b>.0498</b>
	Aktivierung	1248.93	135	.816	<b>.074</b>	<b>.0563</b>
	Klassenführung	1208.54	119	.817	<b>.078</b>	<b>.0545</b>
	Lernförderliches Klima	853.18	77	.877	.081	<b>.0505</b>
	Motivierung	884.13	65	.856	.091	<b>.0587</b>

Das Hinzuziehen der Skaleninterkorrelationen auf latenter Ebene (s. Tabelle 17) zeigt im Einklang mit den bisherigen Ergebnissen geringe Interkorrelationen der Aktivierung mit Klassenführung und dem Lernförderlichen Klima. Dabei orientiert sich diese Einstufung an Brosius (2002), der ab einer Skaleninterkorrelation von .60 von einem hohen Wert spricht. Allerdings weist auch die Aktivierung mit den übrigen Skalen hohe Interkorrelationen auf. Bei den anderen Merkmalen zeigen sich ebenfalls hohe Interkorrelationen. Die größte Korrelation besteht zwischen dem Faktor Klarheit und Verständlichkeit und dem Motivierungsfaktor mit  $r = .829$ . Brosius (2002) rät ab einer Interkorrelation von .60 zum Zusammenfassen der Skalen. Da nur zwei der zehn Interkorrelationen unterhalb dieses Wertes liegen, wird deutlich, dass keine hinreichende diskriminante Validität zwischen den Skalen angenommen werden kann.

Tabelle 17: *Interkorrelationsmatrix der Unterrichtsmerkmalsskalen*

Korrelation	Klassenführung	Lernförderliches Klima	Motivierung	Klarheit und Verständlichkeit	Aktivierung
Klassenführung	1				
Lernförderliches Klima	.793	1			
Motivierung	.678	.697	1		
Klarheit und Verständlichkeit	.744	.811	.829	1	
Aktivierung	<b>.477</b>	<b>.521</b>	.768	.730	1

Anmerkung: Interkorrelationen < .60 sind hervorgehoben, wobei KF = Klassenführung, LK = Lernförderliches Klima, M = Motivierung, KV = Klarheit und Verständlichkeit, A = Aktivierung

Es muss jedoch auch berücksichtigt werden, dass sich die Faktoreninterkorrelation bei einer niedrigen Kommunalität der beobachteten Variablen erhöht (Bühner, 2006) und der Fragebogen nicht den Anspruch erhebt, unabhängige Faktoren zu erfassen.

Wird ein Modell spezifiziert, in welchem die Items, die sich aufgrund der exploratorischen Faktorenanalyse in hoher Übereinstimmung mit den bisherigen konfirmatorischen Analysen als ungeeignet herausgestellt haben (s. Anhang D.3.4) aus dem Modell entfernt werden, zeigt sich allerdings keine deutliche Verbesserung zu der eingangs berichteten Modellgüte des spezifizierten Ausgangsmodells (s. Tabelle 18).

Eine weitere Erklärung für die Befunde könnte der starke Einfluss eines Generalfaktors der Unterrichtseinschätzung bzw. des -klimas sein, die vor dem Hintergrund exploratorischer Analysen naheliegend ist (s. Abschnitt 9.3.2). Dieser wird auf Gesamtmodellebene anhand eines Schmid-Leiman-Modells spezifiziert, das aus den einzelnen Itemgruppen einen Generalfaktor als direkte Ursache für die Varianz-Kovarianzmatrix der manifesten Variablen zieht und zusätzliche latente Gruppenfaktoren mit einem indirekten Effekt ergänzt (Schmid & Leiman, 1957, s. Anhang D.3.5). Die Fit-Indizes in Tabelle 18 zeigen in der Tat eine Verbesserung der Modellgüte durch Hinzunahme eines Generalfaktors sowie latenter Gruppenfaktoren (s. Tabelle 18), die sich allerdings nur auf den RMSEA und SRMR bezieht ( $\chi^2$ -Wert = 9827.68, df = 2837, CFI = .817, RMSEA = .040, SRMR = .041).

Werden die ich- und wir-Version getrennt betrachtet, zeigen sich gute Modellfits. Dies unterstützt die Überlegung, dass der ich- vs. wir-Bezug im Gesamtmodell als erheblicher Methodenvarianzanteil einfließt.

Tabelle 18: *Modell-Fit-Indizes zu verschiedenen Modellvarianten auf der Gesamtebene*

Modellvariante	Modell-Fit-Indizes der konfirmatorischen FA				
	$\chi^2$	df	CFI	RMSEA	SRMR
Modell 1: 5-Faktorenlösung	10454.80	2906	.803	<b>.041</b>	<b>.053</b>
Modell 2: 5-Faktorenlösung (ohne Items, die bei exploratorischer Faktorenanalyse eher ungeeignet)	8748.10	2327	.818	<b>.043</b>	<b>.053</b>
Modell 3: Schmid-Leiman-Lösung	9827.68	2837	.817	<b>.040</b>	<b>.041</b>

Anmerkung: Beim Modell 1 und 2 sind die Fehlerkorrelationen inhaltsäquivalenter Items gefixt.

### 9.3.4 Zusammenfassende Bewertung

Die exploratorische und die konfirmatorische Faktorenanalyse weisen eine hohe Schnittmenge in Bezug darauf auf, welche Items – aufgrund einer zu geringen Varianzaufklärung

durch den Faktor – zu eliminierenden sind. Dazu gehören die Items SKF2, SKF6, SKF10 und SKF12. Die exploratorische Analyse legt zusätzlich die Entfernung der Items SKF15 („Die Klassenregeln waren mir/uns bekannt“) und SA76 („Der Lehrer hat Fragen gestellt, die mich/uns zum Nachdenken angeregt haben“) aufgrund einer Faktorladung von  $\leq .30$  nahe, wobei diese Befunde nicht in Einklang mit den Ergebnissen der konfirmatorischen Faktorenanalysen und den Item- und Reliabilitätsanalysen stehen (s. Abschnitt 9.3.2). Ersteres könnte auf die abweichende Schätzmethodik (exploratorische Faktorenanalyse: HKA, konfirmatorische Faktorenanalyse: ML), letzteres auf unterschiedliche Berechnungsgrundlagen zurückzuführen sein. Denn auch wenn das Item keine Residualvarianz in der konfirmatorischen Analyse aufweist und dadurch zu einem perfekten Modellfit beitragen würde, könnten auf korrelativen Verfahren beruhende Reliabilitätswerte eher gering sein, da diese eine hohen unique Varianz eher bestrafen und hohe Interkorrelationen belohnen.

Bei einem Cut-off-Wert von einem aufgeklärten Varianzanteil kleiner 10% würden gemäß der konfirmatorischen Analyse noch die beiden Items SLK19 und SKV54 hinsichtlich einer negativen Selektion empfohlen. Diese Items weisen auch in der exploratorischen Analyse in der Gesamtstichprobe geringe Faktorladungen auf (SLK 19 =  $-.34$ , SKV54 =  $.30$ ).

Die diskriminante Validität der einzelnen Skalen ist aus statistischen Gesichtspunkten nicht zufriedenstellend. Aus der Anwendungsperspektive heraus ist eine Trennung der einzelnen Skalen jedoch nicht sinnvoll, um für möglichst viele Unterrichtsmerkmale, die im Qualitätsrahmen aufgeführt sind, Indikatoren angeben zu können.

## **9.4 Weitere Validitätsanalysen**

Im Folgenden werden nach einer allgemeinen Darstellung des Vorgehens (Abschnitt 9.4.1), die Inhalts- und Augenscheinvalidität als weitere wichtige Validitätskriterien bezüglich der Itemselektion hinzugezogen (Abschnitt 9.4.2 und 9.4.3). Abschnitt 9.4.4 stellt einen kurzen Exkurs zur Übereinstimmung der Schüler- und Lehrerangaben in Bezug auf die Unterrichtsmerkmale dar.

### **9.4.1 Vorgehen**

Im ersten Schritt wird die Inhaltsvalidität des Fragebogens beurteilt. Da es keinen Kennwert zum Erfassen der Inhaltsvalidität gibt, muss diese „...aufgrund logischer und fachlicher Überlegungen...“ ermittelt und „...mit oder ohne Einschränkung akzeptiert oder



verworfen...“ werden (Michel & Conrad, 1982, S. 57). Murphy und Davidshofer (2001) geben diesbezüglich eine Strukturierungshilfe, indem sie zur Beschreibung des zu erfassenden Konstrukts auf der Inhaltsebene und der anschließenden Festlegung, welcher Bereich durch welche Items abgedeckt ist, raten.

Danach werden wesentliche Befunde zur Augenscheinvalidität fokussiert. Tent und Stelzl (1993) weisen darauf hin, dass Inhalts- und Augenscheinvalidität leicht zu verwechseln sind. Deswegen soll an dieser Stelle eine definitorische Abgrenzung erfolgen. Die Inhaltsvalidität gibt an, „...inwieweit ein Test oder ein Testitem das zu messende Merkmal repräsentativ erfasst“ (Moosbrugger & Kelava, 2007, S. 15), wohingegen die Augenscheinvalidität darauf abzielt, ob die aus den Items ersichtliche Messintention der tatsächlichen Messintention des Tests entspricht (Kersting, 2003; Moosbrugger & Kelava, 2007). Cattell, Eber & Tatsuoka (1970) sprechen daher im Englischen von ‚faith validity‘ statt ‚face validity‘, um das Wecken der subjektiven Annahme beim Probanden hervorzuheben.

#### **9.4.2 Inhaltsvalidität**

Der Fragebogen zielt auf die Erfassung einer Auswahl der von Helmke (2007c) postulierten Merkmale guten Unterrichts ab, die hinsichtlich ihrer Validität in Bezug auf kognitive und z. T. emotional-motivationale Kriterien empirisch fundiert sind (s. Abschnitt 2.2.4). Die deduktive Konstruktionsmethode (Bühner, 2006) erfolgt durch Experten im Bereich der Unterrichtsforschung (s. Kapitel 5) unter Rückgriff auf empirische Befunde langjähriger Forschung zum Konstrukt der Unterrichtsqualität und den Erfahrungen aus einer Reihe von Studien im schulischen Kontext, wie z. B. MARKUS, DESI und VERA. In Bezug auf die fünf im Fragebogen erfassten Konstrukte kann die Inhaltsvalidität daher ohne wesentliche Einschränkungen angenommen werden.

#### **9.4.3 Augenscheinvalidität**

Um Items mit Auffälligkeiten hinsichtlich der Augenscheinvalidität bei den beiden Zielgruppen Schüler und Lehrkräfte zu identifizieren, wurden diese im Anschluss an die durch die Autorin selbst durchgeführten Haupterhebungen im Klassenverband (s. Abschnitt 7.1) zu den folgenden Aspekten befragt: (1) Ob ihnen beim Ausfüllen des Fragebogens bezüglich der Itemauswahl und -formulierung etwas aufgefallen ist und (2) ob es Schwierigkeiten beim Einschätzen der Items gab. Auch die bereits in Abschnitt 7.2 genannte Kontrolle aller Fragebögen hinsichtlich offener Angaben und Streichungen bestimmter Fragen bzw. ver-

merkter Kommentare geben einige augenscheinvaliditätsbezogene Hinweise. In Tabelle 19 sind die qualitativen Angaben zusammengefasst, die von mindestens fünf Klassen und einer Lehrkraft genannt werden.

Tabelle 19: *Qualitative Angaben der Haupterhebung, die hinsichtlich der Augenscheinvalidität auffällig sind*

Item	Itemlabel	Klassenangaben	n	Lehrkraftangabe	n
SKF6	Material oder Geräte waren gleich verfügbar und einsatzbereit	nicht benutzt	8	außer der Tafel keine verwendet	1
SM36	Arbeitsblätter machten Spaß	keine Arbeitsblätter eingesetzt	13	keine Arbeitsblätter verwendet	5
SKV43	Wechsel zwischen Medien hat Stunde interessant gemacht	kein Medienwechsel vorgenommen	8	diesmal keiner erfolgt	1
SKV51	Lehrkraft hat Lerninhalte aufgegriffen, die in den Hausaufgaben vorkamen	keine Hausaufgaben erhalten	10	keine Hausaufgaben (in Ganztagsklassen)	4
SKV56	Lehrkraft achtete darauf, dass Schüler deutsche Sprache richtig verwenden	im Fremdsprachenunterricht nicht zutreffend	21	im Fremdsprachenunterricht nicht zutreffend	2
SA61	Lehrkraft hat Hinweise gegeben, wie Schüler besser lernen können	in Klassenstufen 5-7 gemacht; nicht in dieser Stunde gemacht	8	Information wird regelmäßig, aber in größeren zeitlichen Abständen gegeben	2

Anmerkung: Das n gibt die Anzahl der Klassen/Lehrkräfte an, die diese Aussage getroffen haben.

Außerdem werden durch diese Methode konditionale Items verifiziert. Bei diesen besteht dann eine Antwortambivalenz, wenn eine Teilbedingung des Items nicht erfüllt ist. Das Format von konditionalen Items liegt bei den folgenden Items vor: SKF6, SKF10, SKF12, SLK23, SLK24, SM36, SM42, SM43, SKV51, SKV53, SKV54, SKV56 und SA72. So ist z. B. beim Item SM36 („Die Arbeitsblätter haben mir/uns Spaß gemacht“) unklar, ob ein vom Schüler angekreuztes ‚stimme nicht zu‘ bedeutet, dass keine Arbeitsblätter zum Einsatz kamen oder diese zwar verwendet, aber eben nicht als anregend eingestuft wurden. Die Interviews ergeben, dass die Schüler beim Nichtzutreffen konditionaler Iteminhalte unterschiedliche Beantwortungsstrategien anwenden. Einige lassen die Beantwortung des Items aus, andere nehmen die Einschätzung dennoch vor. Auf die Nachfrage, welche Situation der Beurteilung dann zugrunde liegt, wird auf den Unterricht bei der jeweiligen Lehrkraft im Allgemeinen bzw. frühere Unterrichtssituationen rekurriert. Deshalb wird der Abschluss konditionaler Items, die zu einer deutlichen Antwortambivalenz führen, empfohlen.

#### 9.4.4 Exkurs: Übereinstimmung zwischen Schüler- und Lehrerangaben

Die Übereinstimmung zwischen der Lehrkraft- und Schülerwahrnehmung soll in diesem Abschnitt lediglich kurz skizziert werden, da die Analyse dieser vom Umfang her einer separaten Publikation bedarf. Wesentliche Ergebnisse sollen dem interessierten Leser an dieser Stelle jedoch nicht vorenthalten werden: Die Korrelationen zwischen den Schüler- und Lehrerangaben betragen durchschnittlich  $r = .30$  und reichen von  $r = -.07$  (KV53, „Gelegentliche Zusammenfassungen haben mir geholfen, den Lernstoff besser zu verstehen“) bis  $r = .75$  (KF2, „Niemand kam zu spät“). Hohe Korrelationen zeigen sich v. a. bei Items, die in der Lehrer- und Schülerversion gleich formuliert sind und objektive Sachverhalte erfassen, wie z. B. beim Item KF1 „Die Stunde hat pünktlich begonnen“.

Beim Unterrichtsmerkmal Lernförderliches Klima sind die Übereinstimmungen zwischen der Schüler- und Lehrerwahrnehmung im Vergleich zu den anderen Unterrichtsmerkmalen am größten bzw. die Abweichungswerte am geringsten (Summe der an der Itemzahl gewichteten betragsmäßigen Abweichungswerte  $d^{12} = .08$ ). Die Motivierungsitems zeigen die höchsten Abweichungen ( $d = .25$ ) in dem Sinne, dass die Schüler den Aussagen stärker zustimmen. Eine einfaktorielle ANOVA mit den betragsmäßigen Abweichungswerten  $\left| \bar{x}_{i,Lehrer} - \bar{x}_{i,Schüler} \right|$  pro Item als AV und den Unterrichtsmerkmalsbereichen als UV zeigt allerdings, dass sich die Unterrichtsmerkmale hinsichtlich ihrer Abweichungen auf einem Niveau von  $p < .05$  nicht signifikant unterscheiden ( $F = 2.15$ ,  $df = 4$ ,  $p = .083$ , s. Anhang D.4.1). Die deskriptiv gesehen hohe Abweichung der Motivierungsitems ist allerdings maßgeblich auf das Item M39 („Der Lehrer hat erklärt, dass der Unterrichtsstoff für das Lernen in anderen Fächern wichtig ist“) zurückzuführen. Bei diesem Item weichen die Schüler im Vergleich zu den Lehrkräften fast eine Antwortkategorie nach oben ab. Dies könnte daran liegen, dass die Schüler wissen, was in den anderen Unterrichtsfächern behandelt wurde, die Lehrer allerdings nicht und den Stoff anderer Fächer aufgreifen, ohne dass ihnen dies bewusst ist. Die drittgrößten Abweichungen weist das Merkmal Aktivierung auf ( $d = .21$ ), ebenfalls in der Richtung einer höheren Schülereinschätzung.

---

<sup>12</sup> Der Abweichungswert  $d = \frac{\sum_i^m \left| \bar{x}_{i,Lehrer} - \bar{x}_{i,Schüler} \right|}{m}$ , wobei  $m$  die Summe der Items  $i$  ist.

Die Rangfolge der Zustimmungsmittelwerte der fünf Merkmalsbereiche ist bei Schülern und Lehrern nahezu gleich: Lernförderliches Klima ( $M_{\text{Schüler}} = 3.31$ ,  $M_{\text{Lehrer}} = 3.36$ ), Klassenführung ( $M_{\text{Schüler}} = 3.01$ ,  $M_{\text{Lehrer}} = 3.10$ ), Klarheit und Verständlichkeit ( $M_{\text{Schüler}} = 3.03$ ,  $M_{\text{Lehrer}} = 3.07$ ), Motivierung ( $M_{\text{Schüler}} = 2.67$ ,  $M_{\text{Lehrer}} = 2.51$ ) und Aktivierung ( $M_{\text{Schüler}} = 2.50$ ,  $M_{\text{Lehrer}} = 2.33$ ).

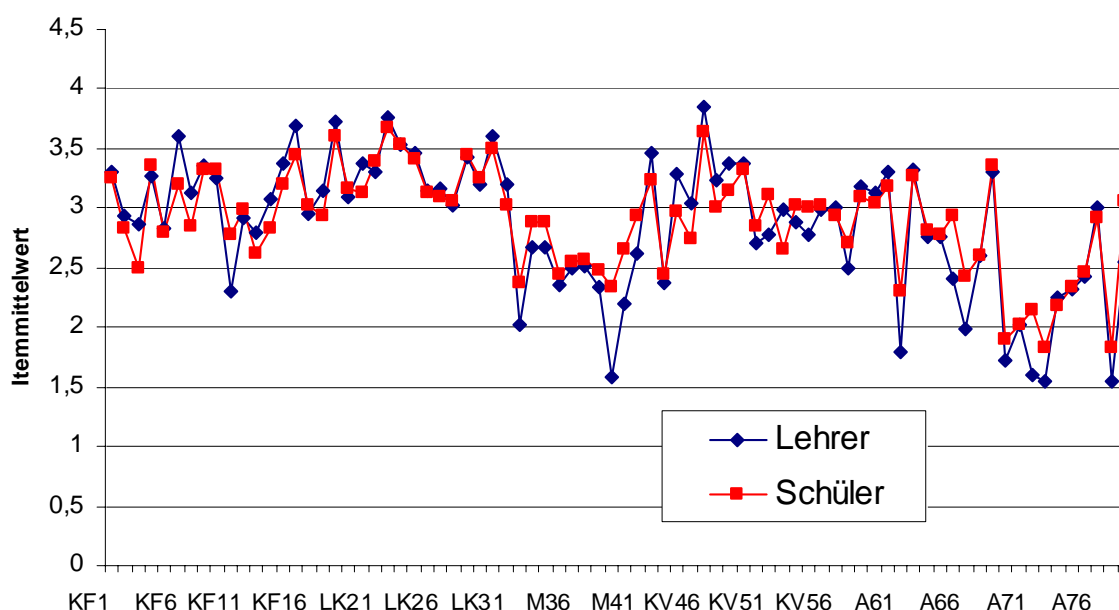


Abbildung 23: Gegenüberstellung der Itemmittelwerte für die Lehrer- vs. Schülerstichprobe

Die Abbildung 23 gibt einen Eindruck der Abweichungen zwischen den Schülern und Lehrern pro Item auf Mittelwertsebene. Nach dem bereits erwähnten Item M39 mit dem höchsten Abweichungswert in Bezug auf eine höhere Einschätzung durch die Schüler, weist das Item KF6 („Material oder Geräte (z. B. Overhead-Projektor) waren gleich verfügbar und einsatzbereit“) auf deskriptiver Ebene die höchste Abweichung im Sinne höherer Zustimmungswerte in der Lehrer- als in der Schülerstichprobe auf. Dies könnte auf eine Wahrnehmungskongruenz zurückzuführen sein. In Folge der Beobachterrolle der Schüler werden Verzögerungen stärker wahrgenommen als aus der agierenden Rolle der Lehrkraft. Die Gesamtbetrachtung der Befunde zeigt aber, dass die Übereinstimmung zwischen den beiden Perspektiven – auf deskriptivstatistischer Ebene betrachtet – relativ hoch ist.

#### 9.4.5 Zusammenfassende Bewertung

Die inhaltliche Validität des Fragebogens ist hinsichtlich des intendierten Ausschnitts aus dem Meta-Konstrukt Unterrichtsqualität als sehr gut zu bewerten. Bezüglich der Augen-

scheinvalidität zeigen sich auffällige Items. Gemessen an der absoluten Häufigkeit der qualitativen Angaben sollten die Items SM36, SKV51 sowie SKV56 unbedingt entfernt werden. Die Items SKF6, SKV43 sowie SA61 zeigen ebenfalls Auffälligkeiten, werden allerdings von einer geringeren Schüleranzahl als problematisch eingestuft.

## 10 Analyse der Angaben zur Urteilssicherheit

Die Variable Urteilssicherheit erfasst die subjektiv empfundene Sicherheit bei der Beurteilung des jeweiligen Items zur Erfassung der Unterrichtsmerkmale. Der Gesamtmittelwert der subjektiven Urteilssicherheit, über beide Itemformulierungen hinweg, beträgt 3.55 bei einem Streubereich von [3.40; 3.82] und weist eine mittlere Streuung von 0.67 mit einem Streubereich von [0.44; 0.81] auf. Wie in Abbildung 24 dargestellt, liegen somit alle Itemmittelwerte in den beiden oberen Antwortkategorien.

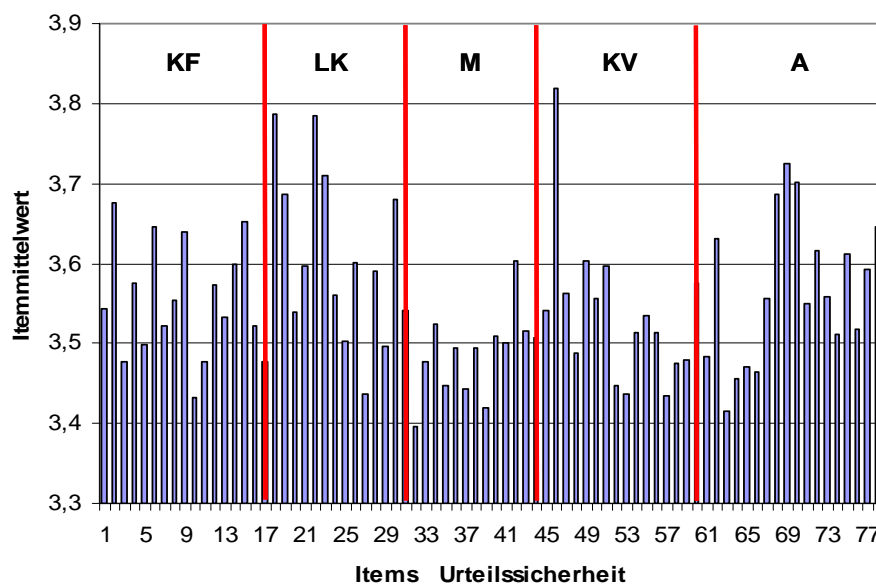


Abbildung 24: Verteilung der Itemmittelwerte der Urteilssicherheit innerhalb der theoretisch postulierten Unterrichtsmerkmalsbereiche (KF = Klassenführung, LK = Lernförderliches Klima, M = Motivierung, KV = Klarheit und Verständlichkeit, A = Aktivierung)

In einer Gegenüberstellung der durchschnittlichen Itemmittelwerte innerhalb der theoretisch postulierten Unterrichtsmerkmale weisen die Items zum Erfassen der Motivierung und, bis auf das Item SKV 46, auch die Items zum Erfassen des Unterrichtsmerkmals Klarheit und Verständlichkeit geringe Mittelwerte in der Urteilssicherheit auf. Die Mittelwerts- und Streuungsanalysen der beiden Fragebogenversionen zeigen auf deskriptiver Ebene nur geringfügige Unterschiede. So beträgt der Mittelwert der ich-Version 3.57 bei einem Streubereich von [3.41; 3.80] und einer durchschnittlichen Standardabweichung von

0.66 bei einem Streubereich von [0.44; 0.81]. Die wir-Version weist einen Mittelwert von 3.54 mit einem Streubereich von [3.38; 3.83] sowie einer durchschnittlichen Standardabweichung von 0.67 mit einem Streubereich von [0.42; 0.81] auf. Abbildung 25 gibt einen Überblick über die Urteilssicherheitsmittelwerte bei den 78 Items in Abhängigkeit von der ich- und wir-Version.

Im Ranking der Itemmittelwerte zeigt sich, dass das Item SM32 („Die Lehrkraft hat Verbindungen zu anderen Fächern hergestellt“) den niedrigsten Mittelwert hat, wobei ‚niedrigste‘ in diesem Fall 3.40 bedeutet und somit deutlich oberhalb des theoretischen Skalenmittels liegt. Danach folgen die Items SA63 („Die Lehrkraft hat uns Tipps gegeben, wie wir Aufgaben selbstständig lösen können“) sowie SM39 („Die Lehrkraft hat erklärt, dass der Unterrichtsstoff für das Lernen in anderen Fächern wichtig ist“).

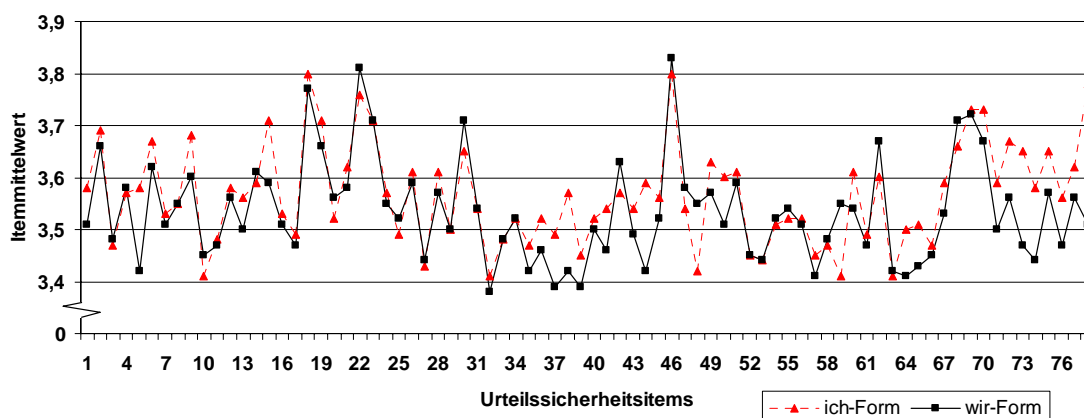


Abbildung 25: Verteilung der Itemmittelwerte in Abhängigkeit von der ich- vs. wir-Version

Den höchsten Mittelwert weist SKV 46 („Die Lehrkraft hat so laut und deutlich gesprochen, dass wir alles sehr gut verstehen konnten“) mit 3.82, den zweithöchsten SLK 18 („Der Umgangston zwischen der Lehrkraft und den Schülern/innen war freundlich“) und den dritthöchsten SLK22 („Die Lehrkraft hat die Klasse zu Beginn der Stunde freundlich begrüßt“) auf.

Die Angaben zur Urteilssicherheit in Bezug auf die einzelnen Items streuen mit einer mittleren Standardabweichung von 0.67, wobei das Minimum des Standardabweichungskennwerts beim Item SKV46 bei 0.44 liegt ( $sd_{\text{ich-Version}}(\text{SLK18}) = 0.44$ ,  $sd_{\text{wir-Version}}(\text{SKV46}) = 0.42$ ) und das Maximum 0.81 beim Item SM32 in allen Versionen. Diese Extrema der Standardabweichungen lassen sich in Abbildung 26 gut erkennen.

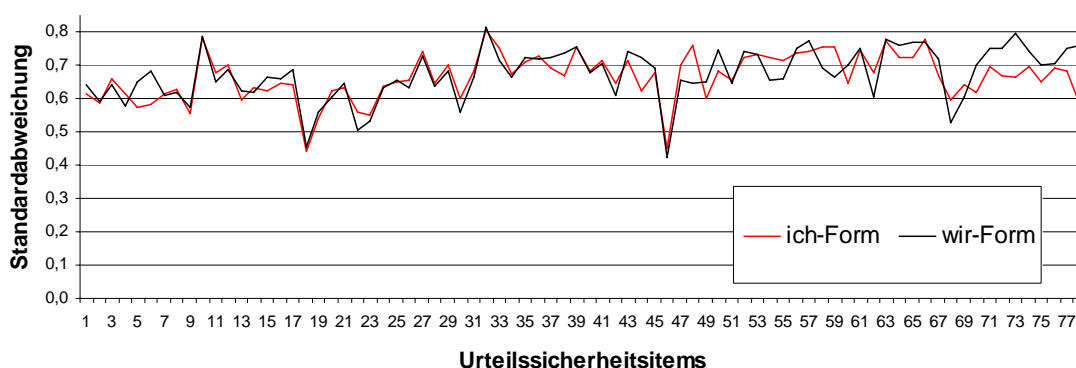


Abbildung 26: *Standardabweichungen der Urteilssicherheitsangaben in Abhängigkeit vom Item und der Fragebogenversion*

Die Verteilungen der Urteilssicherheitsangaben pro Item erweisen sich mit einer durchschnittlichen Schiefe von -1.50 (ich-Version: -1.58; wir-Version: -1.42) als deutlich links-schief – sowohl auf Individual- als auch auf Klassenebene. Modus und Median fallen bei allen Urteilssicherheitsitems mit einem Wert von 4.0 zusammen. Die Differenz aus Median und Mittelwert ist bei allen Items positiv, d. h. die Annahme einer deutlichen Linksschiefe wird auch hiermit bestätigt. Die prozentuale Kategoriehäufigkeit der Antwortalternative ‚sehr unsicher‘ beträgt in der Gesamtstichprobe über alle Items hinweg lediglich 1.49%, wohingegen 64.36% auf die Antwortkategorie ‚sehr sicher‘ entfallen. Die Urteilssicherheitsangaben zeigen somit eindeutige Deckeneffekte. Diese Variable könnte als Selektionskriterium für die Itemauswahl herangezogen werden, in dem Sinne, dass bei geringer subjektiver Sicherheit die Beurteilung des Items schwer fällt und dieses deshalb eliminiert wird. Gründe für eine geringe Urteilssicherheit können ein fehlendes oder unvollständiges Itemverständnis, eine fehlende oder eingeschränkte direkte Beobachtbarkeit des Itemsachverhalts durch die Schüler sowie Erinnerungs- und Motivationsprobleme sein. Ferner könnte die Urteilssicherheit als Gewichtungsfaktor für die Einschätzung der Unterrichtsmerkmale verwendet werden, so dass Items, bei deren Einschätzung die Schüler eher unsicher waren, weniger stark in den Endskalenwert bzw. Bewertung eingehen. Diese Optionen können aufgrund des starken Deckeneffekts jedoch nicht realisiert werden, da diese Variable keinen inkrementellen Informationszuwachs liefern würde. Eine Erklärung in Bezug auf die Deckeneffekte liefert der Befund aus den Interviews im Rahmen der Haupterhebung, dass die Schüler ihre Anwesenheit im Unterricht mit einer hohen Sicherheit gleichsetzen.

## 11 Itemselektion

Im Folgenden sollen die in den vorangegangenen Abschnitten ermittelten Analyseergebnisse hinsichtlich der endgültigen Itemauswahl zusammengeführt werden. Zuerst erfolgt die Bewertung der Ergebnisse zur Eignung der ich- vs. wir-bezogen formulierten Items (Abschnitt 11.1), anschließend die Selektion auf Itemebene (Abschnitt 11.2).

### 11.1 Eignung der ich- vs. wir-Version

Hinsichtlich der Bewertung der Eignung der ich- vs. wir-Version werden die in den vorangegangenen Abschnitten erhobenen Analyseergebnisse in Tabelle 20 gegenübergestellt.

Die deskriptivstatistischen Angaben, die Ergebnisse der Item- und Reliabilitätsanalysen sowie die Mittelwertsangaben der Urteilssicherheit für die beiden Fragebogenversionen erlauben keine eindeutige Entscheidung in Bezug auf die beiden Itemtypen. Das standardisierte Cronbachs  $\alpha$  ist für die wir-Version zwar etwas höher, die Aussagekraft der Kennwerte auf Skalenebene ist allerdings eingeschränkt, da diese auch neutrale Items berücksichtigen, deren Einfluss nicht dem spezifischen ich- bzw. wir-Bezug zugerechnet werden darf. Konfirmatorische Analysen zeigen sowohl im Falle der Modellspezifikation mit wir-Items als auch der mit ich-Items einen guten Modellfit. Die in Abschnitt 3.4.1 beschriebenen zugrunde liegenden Urteilsprozesse sprechen für die Verwendung der ich-Version, v. a. aufgrund potentieller Widersprüche zwischen der Itembeantwortung und dem subjektiven Erleben bei der wir-Version. Ich-bezogen formulierte Items weisen weiterhin den Vorteil auf, dass die Lehrkraft eine Ereignishäufigkeit ermitteln kann. Demzufolge wäre es möglich, die Zustimmungen zu dem Item SA78 („Ich bin während der Unterrichtsstunde mindestens einmal drangekommen“) zu addieren, um zu ermitteln, wie viele Schüler mindestens einmal in der Stunde von der Lehrkraft aufgerufen worden sind. Werden die Fragebögen durch die Lehrkraft selbst ausgewertet (s. Abschnitt 4.3.1), steht diesem Vorteil allerdings die Gefahr der Anonymitätsverletzung gegenüber. So erlauben insbesondere ich-bezogene (und z. T. neutral formulierte) Items die Identifikation der agierenden Person, sofern nur einer oder wenige in der Klasse die im Item erfragte Verhaltensweise gezeigt haben. Die Items SA67 („Ich habe in der Stunde meine eigenen Arbeitsergebnisse vorgestellt“) und SKF12 („Einzelne Schüler/innen haben dem Lehrer geholfen (z. B. beim Austeilen von Arbeitsblättern, dem Abwischen der Tafel oder anderen Dingen)“) ließen z. B. die eindeutige Identifizierung des Schülers unter den genannten Rahmenbedingungen zu.



Tabelle 20: *Ergebnisgegenüberstellung für die ich- vs. wir-Version*

	ich-Version	wir-Version
%ualler Anteil fehlender Werte	0.1-5.8%	0.1-6.1%
M/sd	M = 2.89, sd = 0.98 (ohne neutral formulierte Items: M = 2.88, sd = 0.96)	M = 2.90, sd = 0.98 (ohne neutral formulierte Items: M = 2.91, sd = 0.96)
Trennschärfe	mittlere $r_{itc}$ = .47	mittlere $r_{itc}$ = .48
stand. Cronbachs $\alpha$	Gesamttest: .952	Gesamttest: .956
CFA	$\chi^2/df$ = <b>2.36</b> CFI = .792 RMSEA = <b>.042</b> SRMR = <b>.0566</b>	$\chi^2/df$ = <b>2.40</b> CFI = .797 RMSEA = <b>.043</b> SRMR = <b>.0578</b>
Urteilssicherheit	M = 3.57, sd = 0.66	M = 3.54, sd = 0.67
Fragebogen- bearbeitung	(+) Bezug auf individuelle Unter- richtseinschätzung	(-) Aggregierung auf Klassen- ebene erforderlich; u. U. Wider- spruch zum subjektiven Erleben (z. B. SKF5 „Wir waren die ganze Stunde über aufmerksam und konzentriert“)
Auswertung	(+) bei einigen Items Ereignishäufig- keit ermittelbar (z. B. SA78 „Ich bin mindestens einmal in der Stunde drangekommen“)	
Anonymität	(-) Identifizierung einzelner Schüler möglich (z. B. „Ich habe Lehrkraft durch Bedienung von technischen Geräten unterstützt“ – nur ein Schüler hat in der Stunde geholfen)	
Nebenaspekte	(+) erfordert theoretisch weniger kognitive Verarbeitungskapazität und stärkt die Eigenverantwortung	

*Anmerkung:* Der Modellfit der CFA bezieht sich auf ein Gesamtmodell mit allen 78 Items, bei dem Fehlerkorrelationen zwischen inhaltsgleichen Items und Interkorrelationen der fünf Merkmale zugelassen werden und entweder die ich- oder. wir-Angaben zugrunde gelegt werden.

Die Gewichtung der einzelnen Aspekte im Projektteam bringt letztendlich die Entscheidung der Verwendung ich-bezogen formulierter Items.

## 11.2 Eignung auf Itemebene

Aufgrund der Tatsache, dass die Itemselektion eine Funktion der bei der Fragebogenentwicklung in Betracht gezogenen Gütekriterien ist, die sich mitunter gegenseitig ausschlie-

ßen oder widersprechen können, wurden im mehrstufigen Überprüfungsverfahren nicht nach einzelnen Analyseschritten Items selektiert. Denn zieht man z. B. den Trennschärfe-koeffizienten als alleiniges Entscheidungskriterium bei einer Skala mit vielen inhaltlich gleichen Items heran, würde der Skalenwert maßgeblich durch diese homogenen Items determiniert. Andere Items, die ebenfalls zum Konstrukt gehören, erhalten dadurch eine geringe Trennschärfe und werden somit negativ selektiert. Dies schränkt allerdings die Breite des Konstrukts ein und gefährdet die Inhaltsvalidität. Die Zusammenführung und Gewichtung der einzelnen Ergebnisse erfolgt daher analyseübergreifend in diesem Abschnitt unter der Berücksichtigung der Analyseergebnisse, qualitativer Angaben in Bezug auf die Augenscheinvalidität, der expertenbasierten Konstruktbreite im Sinne einer Inhaltsvalidität und praktischer Gütekriterien. Werden die Analyseergebnisse aller betrachteter Kriterien zusammengeführt, ergibt sich das Bild der Tabelle 21, wobei die grau hinterlegten Items aus dem Fragebogen entfernt werden. Die Abfrage der Urteilssicherheit wird aufgrund der gefundenen Deckeneffekte nicht in den endgültigen Fragebogen aufgenommen.

Aufgrund der hohen Übereinstimmung mit einem Großteil der anderen Analyseergebnisse, werden die Items mit einem aufgeklärten Varianzanteil von kleiner 10% in der konfirmatorischen Faktorenanalyse aus dem Fragebogen entfernt (SKF2, SKF6, SKF10, SKF12, SLK19 und SKV54). Dieser Cut-off-Wert scheint gering, ist aber vor dem Hintergrund sinnvoll, dass der Fragebogen einen Interventionscharakter bezüglich der Definition und des inhaltlichen Spektrums von Unterrichtsqualität hat und von daher eine gewisse inhaltliche Breite beibehalten werden sollte. Knapp über der Cut-off-Grenze liegen noch die Items SA68 („Die Lehrkraft hat Aufgaben gegeben, die ich selbstständig lösen sollte“) und SA78 („Ich bin während der Unterrichtsstunde mindestens einmal drangekommen“), deren Elimination aber mit einer deutlichen Reduktion der inhaltlichen Validität einhergehen würde und daher im Fragebogen verbleiben. Eliminiert werden ferner alle inhaltsähnlichen Items, wobei die Entscheidung dabei im Wesentlichen auf den Trennschärfe-koeffizienten und der Faktorladung beruht (SKF4, SKF8, SKF13, SLK26, SM35 und SKV45). Des Weiteren erfolgt die negative Selektion von Items, die hinsichtlich der Augenscheinvalidität sehr auffällig sind (SM36, SM43, SKV51 und SKV56). SM43 („Der Wechsel zwischen verschiedenen Medien (z. B. Tafel, Poster, CD) hat die Stunde für mich interessant gemacht“) wird ebenfalls entfernt, obwohl es nur hinsichtlich der Augenscheinvalidität als ‚auffällig‘ eingestuft wurde. Letzteres trifft auch auf das Item 61 („Die Lehrkraft hat mir Hinweise gegeben, wie ich besser lernen kann (z. B. zur Arbeitsplatzgestaltung“) zu.

Tabelle 21: Zusammenstellung von auffälligen Items mit den relevanten Gütekriterien

Item	Mis- sing- anzahl	Trenn- schärfe	Selektions- kennwert	IÄ	Faktor- ladung (EFA)	Faktor- ladung (CFA)	KI	Qualitative Angaben
SKF2		gering	gering		< .30	.16		
SKF4				✓				
SKF6		gering	gering		< .30	.29	✓	auffällig
SKF8				✓				
SKF10	91	gering	gering		< .30	.28	✓	
SKF12		gering	gering		< .30	.27	✓	
SKF13				✓				
SLK19		gering	gering	✓		.15		
SLK23							✓	
SLK24							✓	
SLK26				✓				
SM35				✓				
SM36	75						✓	sehr auffällig
SM42							✓	
SM43							✓	auffällig
SKV45				✓				
SKV51			gering				✓	sehr auffällig
SKV53							✓	
SKV54		gering	gering			.31	✓	
SKV56			gering				✓	sehr auffällig
SA61								auffällig
SA64			gering					
SA67			gering					
SA68		gering (wir)	gering			.35		
SA72							✓	
SA75			gering					
SA76					≤ .30			
SA78		gering (ich)			< .30 (ich)	.33		

Anmerkung: Die Einstufung ‚gering‘ entspricht bei der Trennschärfe einem  $r_{itc} < .30$  und beim Selektionskennwert  $SK < .20$ . IÄ = Inhaltsäquivalentes Item, das jeweils den geringeren Trennschärfe- und Selektionskennwert im Itempaar bzw. -drilling hat; KI = Konditionales Item. Grau hinterlegt Items werden bei der Erstellung des endgültigen Fragebogens entfernt.

Dieses verbleibt allerdings aufgrund einer angestrebten Konstruktbreite im Fragebogen. Da die hinsichtlich des Selektionskennwerts als gering eingestuften Items SA64, SA67 und SA75 in Bezug auf weitere Kriterien unauffällig sind, verbleiben diese im Fragebogen. SA76 wird trotz einer niedrigen Ladung in der exploratorischen Faktorenanalyse, aufgrund einer substantiellen Ladung in der konfirmatorischen Faktorenanalyse und keinen vorliegenden Einschränkungen der restlichen Kriterien ebenfalls beibehalten. Das Item SKV53

(„Gelegentliche Zusammenfassungen des Lehrers haben mir geholfen, den Lernstoff besser zu verstehen“) wird aufgrund des konditionalen Charakters entfernt. Dies reduziert die Inhaltsvalidität nicht erheblich, weil das Item SKV40 („Der Lehrer hat am Ende der Stunde die wichtigsten Ergebnisse zusammengefasst“) den Aspekt der Wissenssicherung in Form von Zusammenfassungen erfasst, wenn auch mit einem anderen Frequenzbezug. SLK23 („Der Lehrer hat mich immer ausreden lassen, wenn ich dran war“), SLK24 („Der Lehrer hat meine Vorschläge und Ideen ernst genommen“), SM42 („Bei guten Antworten bin ich von dem Lehrer gelobt worden“) und SA72 („Ich habe den Lehrer durch Bedienung von technischen Geräten unterstützt“) sind ebenfalls konditionale Items, allerdings würde deren Ausschluss die Inhaltsvalidität stark einschränken. Die Wahrscheinlichkeit eines Nichtzutreffens der Teilaussage und somit die Auswirkung der negativen Effekte der Konditionalität ist bei drei der Items in der wir-Formulierung reduziert. Deshalb werden diese in der wir-bezogenen Form in den endgültigen Fragebogen als „Der Lehrer hat Vorschläge und Ideen der Schüler/innen ernst genommen“, „Der Lehrer hat die Schüler/innen immer ausreden lassen, wenn sie dran waren“ und „Bei guten Antworten sind die Schüler/innen von dem Lehrer gelobt worden“ aufgenommen. Das Item SA72 weist auch in der wir-Formulierung einen konditionalen Charakter („Die Schüler/innen haben den Lehrer bei der Bedienung technischer Geräte unterstützt“) auf. Außerdem hat es ein relativ hohes Potential der Identifikation der entsprechenden Schüler, was der praktischen Notwendigkeit der Anonymitätswahrung zuwider läuft. Deshalb wird dieses Item ebenfalls entfernt.

Insgesamt werden somit 18 Items aus dem Fragebogen entfernt, so dass der Fragebogen nun 60 Items aufweist, die sich wie folgt auf die Unterrichtsmerkmale verteilen: Klassenführung (10 Items), Lernförderliches Klima (12 Items), Motivierung (10 Items), Klarheit und Verständlichkeit (11 Items) und Aktivierung (17 Items). Der endgültige Fragebogen wird nach der Itemselektion in zwei Versionen, mit einer männlichen und weiblichen Lehrkraftbezeichnung erstellt. Die Version der männlichen Lehrkraftbezeichnung kann im Anhang E.1 eingesehen werden.

Aufgrund des Fokus der Unterrichtsentwicklung auf die Itemebene, der Tatsache, dass Skalenreliabilitäten nicht anhand der Stichprobe ermittelt werden sollten, die zur Konstruktion des Fragebogens herangezogen wurde und des Befunds der hohen Interkorrelation der Skalen, wird auf eine Angabe der Reliabilitätswerte des endgültigen Fragebogens an dieser Stelle verzichtet. Um hinsichtlich einer Manualerstellung skalenspezifische Cronbachs  $\alpha$ -

Werte als Orientierungsgrößen angeben zu können, ist gemäß einem konservativen test-theoretischen Vorgehen die Erhebung einer unabhängigen Stichprobe aus dem Geltungsbereich des Fragebogens indiziert.

Abschließend ein kurzer Ausblick auf die geplanten Veröffentlichungsschritte: Der Fragebogen wird ab Januar 2009 Lehrkräften zusammen mit einem Auswertungstool zur Erstellung des im Abschnitt 7.4 beschriebenen Ergebnisberichts zur Verfügung gestellt. Um die Lernfähigkeit des Instruments zu erhalten, werden des Weiteren ein Feedbackformular für Anmerkungen und Verbesserungsvorschläge und die Möglichkeit zum Datenaupload implementiert, um längerfristig auf Grundlage dieser Referenzwerte für die netzbasierte Stichprobe ermitteln zu können.

## 12 Diskussion

Ausgangspunkt der vorliegenden Arbeit waren Überlegungen zum Mangel an empirisch überprüften schulinternen Feedbackinstrumenten für Lehrkräfte. Aus der Schlüsselrolle des Unterrichts im komplexen schulischen Wirkgefüge leitete sich die Operationalisierung der Unterrichtsqualität als zu erfassendes Konstrukt für ein Selbstevaluationsinstrument ab. Der Fokus lag dabei auf einem fächerübergreifenden verhaltensbasierten Fragebogen, der praktischen Anforderungen im Schulalltag und der Beurteilung aus mehreren Perspektiven gerecht werden sollte. Die Schülerperspektive wurde als Untersuchungsfokus aufgrund des hohen, aber noch weitgehend ungenutzten Potentials von Schülern als Feedbackgeber zu Lehr-Lernprozessen gewählt.

Ausgehend von einem Itempool zu fünf Unterrichtsmerkmalen wurde dazu ein Fragebogen generiert und, im Anschluss an eine qualitative Voruntersuchung, in 80 Klassen weiterführender Schulen durch Schüler und Lehrkräfte der Sekundarstufe I erprobt. Nach der Rückmeldung relevanter Ergebnisse an die teilnehmenden Lehrkräfte wurden die Schülerangaben ausführlichen Analysen unterzogen – mit dem Ziel, diejenigen Items auszuwählen, die sowohl testtheoretischen als auch praktischen Kriterien genügen. Das mehrstufige Itemüberprüfungsverfahren mündete schließlich in einen Fragebogen mit 60 Items zur Erfassung der Unterrichtsqualität auf Prozessebene, der Lehrkräften in Zukunft als Selbstevaluationsinstrument zur Verfügung gestellt werden soll. Für den Einsatz des Fragebogens im schulischen Alltag wurden ein Ablaufmodell zum Schülerfeedback sowie weitere Informationsmaterialien entwickelt.

In diesem Kapitel werden, nach der Reflexion methodischer Elemente dieser Arbeit (Abschnitt 12.1), der endgültige Fragebogen einer kritischen Bewertung unterzogen (Abschnitt 12.2) und weiterführende Forschungs- und Anwendungsfragen aufgezeigt (Abschnitt 12.3). Die Diplomarbeit schließt mit einem Resümee und Ausblick in Bezug auf die Fragestellung der vorliegenden Arbeit (Abschnitt 12.4).

### 12.1 Methodische Überlegungen

*Stichprobe.* Die Schüler- und Lehrerstichprobe der Haupterhebung können hinsichtlich der Datenqualität in Bezug auf fehlende Werte als sehr gut eingestuft werden. Der Umfang der Schülerstichprobe übersteigt auf der Individualebene deutlich den notwendigen Umfang zur

Überprüfung des Fragebogens. Die Lehrkraft- bzw. Klassenanzahl ist im Hinblick auf weiterführende mehrbenenanalytische Verfahren akzeptabel. Während der Erhebung wurde eine Gleichverteilung der Stichprobenelemente im Hinblick auf mehrere Kriterien angestrebt. Diese konnte für das Geschlecht exakt und in Bezug auf die Schulart und die Klassenstufe annähernd erreicht werden. Die Schulart der Dualen Oberschule wurde im Nachhinein ausgeschlossen, da keine Lehrkräfte akquiriert werden konnten. Hinsichtlich der Repräsentativität gibt es eine weitere Einschränkung: Die Befragten entstammen einer anfallenden Stichprobe. Diese bringt den Nachteil der Selbstselektivität insofern mit sich, dass insbesondere selbstkritische engagierte Lehrkräfte an der Befragung teilnehmen, die Feedback als eine essentielle Methode hinsichtlich einer eigenmotivierten Unterrichtsentwicklung ansehen. Diese Kritik bezieht sich in erster Linie auf die Lehrerstichprobe. Sie könnte aber, über die Unterrichtsqualität bzw. das Lehrerverhalten im Allgemeinen mediiert, auch für die Schülerstichprobe gelten – in dem Sinne, dass Klassen überrepräsentiert sind, die hinsichtlich ihrer Aktivität und Aufgeschlossenheit gegenüber Neuem eher im oberen Teil der Normalverteilung liegen. Die Überrepräsentation des positiven Bereichs, die sich methodisch in einer Varianzeinschränkung, d. h. einem *restriction of range* äußert (Wittmann, 1990), kann zu einer Faktorenstruktur führen, die für die Gesamtpopulation nicht repräsentativ ist. Zum Teil konnte die Autorin diesem Effekt durch die zufallsbasierte Akquise weiterer Lehrkräfte vor Ort entgegenwirken, als nivelliert kann die eingeschränkte Repräsentativität dadurch jedoch in keinem Fall angesehen werden. Es bleibt aber auch festzuhalten, dass es sich bei der Stichprobe um keine Normierungs- bzw. Eichstichprobe handelt, bei der eine mangelnde Stichprobenrepräsentativität ein stärkeres Gewicht hinsichtlich potentieller Konsequenzen und der Interpretierbarkeit der Analyseergebnisse hat. Vor dem Hintergrund des praktischen Einsatzzwecks des Verfahrens bleibt auch festzuhalten, dass eine per Zufallsauswahl ‚verordnete‘ Teilnahme an der Befragungsstudie dem Grundsatz eines freiwilligen Schülerfeedbacks zuwider läuft. Außerdem nehmen die an der Befragung teilnehmenden Lehrkräfte gerade wegen ihres Engagements eine wichtige Sekundärfunktion als Multiplikatoren in Bezug auf die Etablierung des Fragebogens ein. Hinsichtlich der qualitativen Voruntersuchung ist ergänzend zu erwähnen, dass die Lehrerstichprobe mit zwei Elementen deutlich zu gering ausfiel. Diese Überprüfung stand jedoch aufgrund des Fokus auf den Schülerfragebogen nicht im Mittelpunkt der vorliegenden Untersuchung.

*Auswertungsmethoden.* Im Rahmen der mehrstufigen Itemüberprüfung wurden testtheoretische Überprüfungen durchgeführt. Die vorliegende Arbeit sieht sich mit der Kritik konfrontiert, dass diese nur bedingt auf den Kontext der Unterrichtswahrnehmung übertragbar sind, weil der Fragebogen als internes Evaluationsinstrument und nicht zur Differenzierung von Schülern mit hohen vs. niedrigen Eigenschafts- bzw. Fähigkeitsausprägungen im Trait-Kontext verwendet werden soll, in dem die Klassische Testtheorie entwickelt wurde. Die statistischen Zusammenhänge, z. B. zwischen der Trennschärfe, der Itemschwierigkeit und der Itemstreuung gelten allerdings auch im vorliegenden Kontext, da laut Bühner (2006) die Klassische Testtheorie eine reine Messfehlertheorie ist und gemäß Borsboom und Mellenbergh (2002) nicht nur Fähigkeiten wahre Werte darstellen müssen. Aufgrund der grundsätzlichen Kritik an der Klassischen Testtheorie (Amelang & Zielinski, 2002; Bühner, 2006; Conrad, 2003) ist der Rückgriff, z. B. auf die Probabilistische Testtheorie, durchaus indiziert und stellt die Möglichkeit einer ergänzenden Überprüfung des Fragebogens dar.

Ferner ist kritisch anzumerken, dass sowohl exploratorische als auch konfirmatorische Faktorenanalysen anhand desselben Datensatzes durchgeführt wurden. Dies ist einem streng wissenschaftlichen Vorgehen folgend nicht zulässig, aufgrund der explorativen Konstruktionsausrichtung der Analysen im vorliegenden Kontext jedoch vertretbar. Eine Teilung der Stichprobe, um die Ergebnisse der exploratorischen Analysen in der anderen Stichprobenhälfte einer konfirmatorischen Kreuzvalidierung unterziehen zu können, erwies sich aufgrund der zu geringen Fallzahl auf Klassenebene für die ich- vs. wir-Version als nicht umsetzbar. Die Möglichkeit, dass bei Verwendung derselben Stichprobe als Grundlage für beide Analysezugänge Ergebnisunterschiede Artefakte der Schätzmethode sein könnten, wurde durch den Fokus auf die Verwendung einheitlicher Verfahren berücksichtigt. Ergänzend wurden Sensitivitätsanalysen durchgeführt, um die Robustheit der Ergebnisse bei Variation der zugrunde liegenden Berechnungsverfahren zu prüfen.

*Analyseebene.* Als eine Schwäche dieser Untersuchung kann angesehen werden, dass die geforderte Mehrebenenanalyse nicht angewendet wurde. Der vorliegende Datensatz ist hierarchisch in die Ebenen Schüler, Klassen, Schulen und Schularten gegliedert. Dieser Umstand wird auch als Clustering bezeichnet (Cohen, Cohen, West & Aiken, 2003). Ein vorliegendes Clustering erfordert eine Mehrebenenanalyse (Ditton, 1998), welche die Effekte der einzelnen Ebenen adäquat berücksichtigt. Diese wurde jedoch nicht angewendet, stattdessen wurde auf die Ansätze einer dis- und aggregierten Analyse zur Überprüfung



der faktoriellen Validitäten auf Individual- und Klassenebene zurückgegriffen. Die einzelnen Datenpunkte sind allerdings nicht unabhängig voneinander, weil sie sich innerhalb der jeweiligen Gruppen bezüglich verschiedener Kriterien (z. B. Sozialisierung, Leistungsniveau etc.) durch eine gemeinsame Lerngeschichte und Selektionsprozesse ähneln. Statistische Standardverfahren, wie sie auch in der vorliegenden Arbeit durchgeführt wurden, setzen diese Unabhängigkeit allerdings voraus (Cohen et al., 2003). Die fehlende Mehrebenenmodellierung kann auf der Individualebene zu fälschlich kleinen Schätzungen des Standardfehlers sowie zu großen Teststatistiken führen (Ditton, 1998). Der adäquaten Herangehensweise einer Mehrebenenmodellierung wurde in der Arbeit aufgrund der Komplexität des Datensatzes in Hinsicht auf das Vorliegen zweier Itemtypen (ich- vs. wir-Bezug) mit jeweils 78 Variablen aus zwei Perspektiven (Lehrer- vs. Schülerstichprobe) und den zusätzlich vorhandenen Urteilssicherheitsangaben im Hinblick auf die Analyse verschiedener Kennwerte nicht entsprochen.

*Überprüfungsversion des Fragebogens.* Die Überprüfungsversion enthielt einige Paare gleichen Inhalts mit dem Hintergrund des Vergleichs der Itemgüte. Diese hätten verschiedenen Testhälften bzw. -versionen zugeordnet werden müssen, denn die Aufnahme in *einen* Test kann (1) zu einer artifiziellen Erhöhung der Skalenkennwerte und einer hohen Fehlerinterkorrelation führen, die wiederum zur Unterschätzung der auf Basis von Cronbachs  $\alpha$  ermittelten Reliabilitätskoeffizienten führen kann. (2) Hinsichtlich der Antwortprozesse kann die Konstellation der Verwendung inhaltsähnlicher Items zu Sequenzeffekten führen und dadurch die Konstruktvalidität des Fragebogens gefährden (Schwartz & Sudman, 1996). Ferner könnte die Konditionalität der Items zu Problemen bei der Fragebogenbearbeitung beigetragen haben, die sich in einer Antwortambivalenz und dadurch erschwerten Interpretierbarkeit der Angaben äußert.

In Bezug auf die Kennwertgüte spielt auch die Itemanzahl eine Rolle. Die Überprüfungsversion verfügt über eine hohe Anzahl von Items. Die durchweg positiven Interkorrelationen dieser 78 Items tragen zur Erhöhung von Cronbachs  $\alpha$  bei, auch wenn sie inhaltlich nicht optimal zur Skala passen (Bühner, 2006). Die ermittelten Cronbachs  $\alpha$ -Werte sind somit eher als Obergrenze der Reliabilität zu interpretieren.

Des Weiteren ist der Aspekt, dass die beiden Fragebogenversionen neben den ich- vs. wir-bezogen formulierten Items auch Items in ‚neutraler‘ Formulierung enthielten, kritisch zu beurteilen. Der Vergleich zwischen der ich- und wir-Version entspricht somit nicht dem

Vergleich zwischen ich-bezogen und wir-bezogen formulierten Items, denn die neutralen Items können zu additiven bzw. Interaktionseffekten in Bezug auf die Schülerangaben führen. Die Analyseergebnisse der ich- und wir-Version sind somit nicht eindeutig dem Effekt eines ich- bzw. wir-Bezugs zuordenbar. Dies wurde in den Analysen an den gegebenen Stellen durch eine separate Betrachtung der ich- bzw. wir-bezogen formulierten und der neutral formulierten Items beachtet.

## 12.2 Kritische Bewertung des Fragebogens

Durch den Fokus des Fragebogens auf die Prozessebene des Unterrichts wird Lehrkräften eine Facette des allgemeinen pädagogischen Wissens und Könnens (Baumert & Kunter, 2006) im Sinne leistungs- und motivationsrelevanter Indikatoren der Unterrichtsführung und „...Orchestrierung von Lerngelegenheiten...“ (S. 487) vermittelt. Darüber hinaus erfasst das Fragebogenverfahren wesentliche Facetten der Aktivierung, Motivierung, Klarheit und Verständlichkeit sowie des Lernförderlichen Klimas im Rahmen des Unterrichts. Im Folgenden soll der Fragebogen in Anlehnung an Bühner (2006) hinsichtlich der psychometrischen Hauptgütekriterien (1) Objektivität, (2) Reliabilität und (3) Validität sowie der Nebengütekriterien (4) Ökonomie und (5) Nützlichkeit bewertet werden.

(1) Die Objektivität als Ausmaß, in dem „die interindividuellen Unterschiede zwischen Probanden unabhängig sind von der Art der Durchführung, Auswertung und Interpretation des Tests“ (Conrad, 2003, S. 24), kann aufgrund der Standardisierung des Fragebogenverfahrens im Allgemeinen als erfüllt angesehen werden (vgl. auch Wagner, 2008).

(2) Die Reliabilität bezieht sich auf die Messgenauigkeit mit dem das untersuchte Konstrukt erfasst wird und stellt eine Voraussetzung für die Validität dar (Amelang & Zielinski, 2002). Im vorliegenden Anwendungskontext ist die Reliabilität eine Funktion instrumenteller Eigenschaften, der situativen Befragungsbedingungen und der Objektivität. Äußere Rahmenbedingungen, wie z. B. die Erhebung zu Beginn vs. am Ende eines Schulalltages, geringe Störgeräusche oder eine ausreichende Beleuchtung während der Befragung lassen sich durch eine adäquate Planung der Befragung teilweise (s. Abschnitt 4.3.1), innere Bedingungen, wie z. B. die Motivation oder Stimmung des Schülers sowie Interaktionseffekte dieser beiden Variablengruppen nur bedingt beeinflussen. Die Überprüfungsversion des Fragebogens zeigt mittlere Cronbachs  $\alpha$ -Reliabilitätswerte der inneren Konsistenz zwischen  $.80 \leq r_{tt} \leq .90$ . In Bezug auf den endgültigen Fragebogen steht die Reliabilitäts-

überprüfung, anhand einer unabhängig von der Konstruktion erhobenen Stichprobe, noch aus und wird als nächster Schritt zur Validierung des Fragebogens empfohlen.

(3) Die Validität bezeichnet gemäß Conrad (2003) den „...Grad der Genauigkeit, mit dem ein Test das misst, was er messen soll, bzw. was er zu messen vorgibt“ (S. 40). Im Kontext von Schülerurteilen unterscheidet Hofer (1982) diesbezüglich zwischen der Übereinstimmung mit inneren und äußeren Kriterien. Eine Übereinstimmung mit inneren Kriterien ist gemäß Hofer (1981) dann gegeben, wenn die „...innere subjektive Realität bei den Schülern angemessen und unverfälscht...“ (S. 53) wiedergegeben werden kann. Dazu müssen eine Reihe von Konstruktions- und Durchführungsbedingungen erfüllt sein. In Bezug auf die Konstruktion liegt bei dem endgültigen Fragebogen ein ausreichend differenziertes, aber nicht zu komplexes Antwortformat vor. Die Modifikationen aufgrund der Ergebnisse der qualitativen Vortests schließen ferner gravierende Probleme hinsichtlich des Itemverständnisses aus. Eine Durchführung des Fragebogenverfahrens entsprechend dem in der vorliegenden Arbeit aufgezeigten Rahmenmodell wahrt die Anonymität der Schüler als Feedbackgeber (s. Abschnitt 4.3.1). Die qualitativen Pretests zeigen, dass ein zusätzliches offenes Antwortformat von einigen Schülern gewünscht wird und deshalb zur Erhöhung der inneren Validität im Sinne von Hofer (1981) beitragen könnte. Eine praktikable Möglichkeit zur Realisierung dieses Formats unter Wahrung der Anonymität ist ein online-basiertes Verfahren. Ein solches ist in Bezug auf den überprüften Fragebogen seitens des IFB in Planung. In der praktischen Bearbeitung könnten den Befragten die wenigen, noch in dem Fragebogen vorliegenden, konditionalen Items Schwierigkeiten hinsichtlich der Entscheidung für eine der Antwortalternativen bereiten. Lösungsmöglichkeiten in Bezug auf den Umgang mit diesen wären: (1) Das Einfügen eines zusätzlichen Feldes im Fragebogen, für diese Unterrichtsstunde keine Angabe möglich'. (2) Die Instruktion der Schüler durch die Lehrkraft bezüglich des Umgangs mit nicht zutreffenden Items, insofern, dass diese ausgelassen bzw. mit Bezug auf den Unterricht bei der Lehrkraft im Allgemeinen beantwortet werden sollen. Gegen die erste Option spricht, dass sie die Schüler dazu verleiten könnte, auch bei nicht-konditionalen Items ein Nichtzutreffen anzugeben. Und die zweite Option hat den Nachteil, dass die Schüler auch bei nicht-konditionalen Items mit dem allgemeinen statt dem konkreten Bezug auf die einzuschätzende Unterrichtsstunde urteilen. Die Vor- und Nachteile müssen für den jeweiligen Anwendungsfall gegeneinander abgewogen werden.

Die Beantwortung der Frage nach der Gültigkeit in Bezug auf äußere Kriterien erfordert empirische Befunde zur Kriteriumsvalidität, z. B. in Form konkurrender und prädiktiver Validitätskoeffizienten. Neben der in deskriptiven Ansätzen aufgezeigten hohen Übereinstimmung der Schüler- und Lehrerangaben (Abschnitt 9.4.4) liegen in Bezug auf den Fragebogen keine Befunde vor. Für die erfassten Unterrichtsmerkmale lässt sich zwar eine Prädiktion kognitiver und emotional-motivationaler Kriterien nachweisen (s. Abschnitt 2.2.4), allerdings bedarf es einer Replikation der Befunde für den endgültigen Fragebogen. Eine kriterienbezogene Validierung ist daher dringend angeraten – auch in Bezug auf die Vorhersage von Kriterien der Unterrichtsentwicklung, operationalisiert über die im Abschnitt 4.1 aufgezeigten Primär- und Sekundärfunktionen des Schülerfeedbacks als Selbst-evaluationsverfahren.

Neben der Übereinstimmung mit der Lehrkrafteinschätzung können weitere Kriterien zur Ermittlung konkurrender Validitätskoeffizienten hinzugezogen werden, wie z. B. die Einschätzung externer Beobachter bzw. die sich aus einer Videografie des Unterrichts ergebenden Daten. Dies kann anhand eines Multitrait-Multimethod-Ansatzes (Campbell & Fiske, 1959) realisiert werden, wodurch Aussagen über die konvergente und diskriminante Validität des Fragebogens möglich werden.

Zur Beurteilung der Konstruktvalidität kann die faktorielle Validität herangezogen werden. Der Fragebogen würde dann eine faktorielle Validität aufweisen, wenn die empirisch beobachtete faktorielle Struktur dem postulierten hypothetischen Konstrukt entspricht (Conrad, 2003). Die 5-Faktorenstruktur konnte in konfirmatorischen Analysen mit einem akzeptablen Modellfit gezeigt werden, wobei die fünf erfassten Merkmale relativ hohe Interkorrelationen zeigen. Relativierend bleibt an dieser Stelle allerdings festzuhalten, dass der Fragebogen nicht den Anspruch erhebt, voneinander unabhängige Faktoren zu erfassen. Bezüglich der Faktorenstruktur ist auf Itemebene positiv hervorzuheben, dass alle Items des Fragebogens eine einheitliche Polung aufweisen. Abwechselnd positiv bzw. negativ gepolte Items können zu artifiziellen Faktorenstrukturen führen, wenn sich die Polung in einem ‚Methodenfaktor‘ abbildet (vgl. Moosbrugger & Kelava, 2007). Gründe praktischer Natur legen ebenfalls das Beibehalten einer bestimmten Polung nahe: Eine einheitliche Polung reduziert die Verarbeitungskomplexität für die Befragten (Angleitner & Riemann, 1996), ebenso die Fehleranfälligkeit bei der Auswertung und Interpretation durch die Lehrkräfte.

Die Inhaltsvalidität kann hinsichtlich der fünf erfassten Unterrichtsmerkmale als sehr gut eingeschätzt werden. Wird allerdings auf alle zehn Merkmale guten Unterrichts rekurriert, müssen bereits insofern Abstriche gemacht werden, dass z. B. Aspekte der Passung, des Umgangs mit der Heterogenität, der Methodenvielfalt oder der Sicherung des Wissens nicht erfasst werden. Zieht man das Verständnis der Klimaforschung heran, fehlt z. B. das Sozialklima im Sinne einer Schüler-Schüler-Orientierung als Teilfacette des Unterrichtsklimas. Darüber hinaus verwendet der Fragebogen fast ausschließlich niedrig-inferente, an konkreten Verhaltensbeispielen verankerbare Itemaussagen. Hinsichtlich der Reliabilität ist dies vorteilhaft, aber dieser Itemtyp bildet durch die Konzentration auf Verhaltenssequenzen hoher Frequenz, aber relativ geringer Zeitdauer, wie z. B. auf Aspekte des Klassenmanagements oder inhalts- bzw. aufgabenbezogener Tätigkeitsanteile (vgl. Clausen et al., 2003) nur einen Teil des möglichen Lehrerverhaltens ab. Die Inhaltsvalidität hat insofern praktische Relevanz, als dass der Fragebogen in seiner Anwendung als Selbstevaluationsinstrument immer auch eine Intervention dahingehend darstellt, dass den Evaluatoren vermittelt wird, durch welche Facetten sich ein ‚guter Unterricht‘ auszeichnet. Dieser Tragweite und der damit eventuell einhergehenden Stärkung des lehrerzentrierten Unterrichts kann durch eine Erweiterung des Fragebogens um oben genannte Aspekte als Zusatzmodule Rechnung getragen werden. Dabei ist anzumerken, dass die Auswahl der erfassten Unterrichtsmerkmale auch den beschränkten zeitlichen Ressourcen im Schulkontext Rechnung trägt.

Bezüglich der angestrebten Augenscheinvalidität seitens der Schüler und Lehrkräfte, könnte der methodische Einwand eines damit einhergehenden Akzeptanz-Verfälschbarkeits-Dilemmas nach Kersting (2003) ins Feld geführt werden. Im Kontext einer systematischen und nachhaltigen Unterrichtsentwicklung (s. Abschnitt 4.3.1) wird dieses Dilemma allerdings zugunsten der höheren Akzeptanz des Verfahrens aufgelöst.

(4) Gemäß Bühner (2006) ist ein Testverfahren dann ökonomisch, wenn es (1) nur eine kurze Durchführungszeit beansprucht, (2) der Materialverbrauch gering ist, (3) die Handhabung einfach ist, (4) das Verfahren in einem Gruppensetting durchführbar sowie (5) schnell und bequem auszuwerten ist. Bis auf den zweiten Aspekt können alle Kriterien als hinreichend erfüllt angesehen werden: Die sich in der Überprüfungsphase ergebende Bearbeitungszeit betrug durchschnittlich 15 Minuten und konnte hinsichtlich der Bearbeitung des endgültigen Fragebogens aufgrund der Itemreduktion und des Entferns der Variable Urteilssicherheit weiter reduziert werden. Der Fragebogen wird Lehrkräften

hinsichtlich der einfachen Handhabbarkeit mit einer Instruktion und einem Rahmenmodell zur Durchführung zur Verfügung gestellt. Die Erhebung im Klassenkontext ist in der Paper-Pencil-Version möglich und bezüglich der Anonymität erforderlich. Aufgrund der Kombination mit einer Excel-Auswertungsmaske ist die Datenauswertung relativ einfach und praktikabel. Das zweite Kriterium kann als bedingt erfüllt angesehen werden, denn die Paper-Pencil-Version umfasst lediglich vier Seiten, würde allerdings auf Klassenebene hochgerechnet, einen gewissen Materialaufwand mit sich bringen. Die onlinebasierte Erhebung wäre in Bezug auf die Materialreduzierung eine Lösung.

(5) Die Nützlichkeit eines Verfahrens kann dann angenommen werden, wenn es „...eine Verhaltensweise misst oder vorhersagt, für...deren Untersuchung ein praktisches Bedürfnis besteht“ (Bühner, 2006, S. 44). Da überprüfte fächerübergreifende Selbstevaluationsinstrumente zur Erfassung der Unterrichtsqualität aus der Schülerperspektive gemäß dem Wissensstand der Autorin nicht vorliegen, Schulen allerdings zur Durchführung interner Evaluationsmaßnahmen gesetzlich verpflichtet sind (s. Abschnitt 1.1), wird dieses Kriterium deutlich erfüllt.

### **12.3 Weiterführende Forschungs- und Anwendungsfragen**

Im Folgenden sollen weiterführende Forschungs- und Anwendungsfragen beschrieben werden, die sich aus den theoretischen Überlegungen und den empirischen Befunden der vorliegenden Arbeit ergeben haben. Diese sind anhand des vorliegenden Datensatzes oder anhand der Erhebung einer neuen Stichprobe realisierbar.

Erstens ist es wichtig, anhand des *vorhandenen Datensatzes* die in dieser Arbeit gefundenen Effekte hinsichtlich der Replizierbarkeit bei Anwendung einer Mehrebenenanalyse zu prüfen, um die Effekte auf Klassen- und Schülerebene getrennt untersuchen zu können. Es ist denkbar, dass sich die Faktorenstruktur in Abhängigkeit von der jeweiligen Ebene unterscheidet.

Zweitens sind Untersuchungen zur faktoriellen Validität des Lehrerfragebogens möglich und nötig, denn der Schülerfragebogen wird als Tandemverfahren in Kombination mit dem Lehrerfragebogen zur Selbsteinschätzung kommuniziert. Die in der vorliegenden Arbeit aufgeführte analytische Entscheidungsgrundlage hinsichtlich der Itemselektion beruht allerdings nur auf den Schülerangaben. Somit lassen sich keine Aussagen über die Güte des Lehrerfragebogens treffen. Nicht in der vorliegenden Arbeit berichtete, aber durchgeführte

exploratorische Faktorenanalysen der Lehrerangaben weisen vergleichbare Tendenzen auf, bedürfen allerdings differenzierter Analysen.

In Anlehnung an den Befund von Clausen (2002), dass verschiedene Beurteilungsperspektiven spezifisch valide sind, kann drittens auch die Übereinstimmung zwischen den Lehrer- und Schülerangaben hinsichtlich der Unterrichtswahrnehmung Gegenstand differenzierter Analysen sein. Eine sich aus den deskriptiven Befunden ergebende Hypothese ist, dass Aktivierungs- und Motivierungselemente eher von einem modernen Bild des selbstständigen Schülers als Adressat und Mitgestalter des Unterrichts zeugen und entsprechende Verhaltensweisen seitens der Lehrkraft von Schülern im Vergleich zur Lehrkraft aufgrund der starken Schülerbezogenheit intensiver wahrgenommen werden. Lehrkräfte könnten motivierende und aktivierende Instruktionsformen als Komponenten der Unterrichtsqualität im Gegensatz zu eher objektiv definierbaren Aspekten der Strukturierung und Klassenführung weniger stark gewichten. Diese Hypothese erfordert entsprechende Untersuchungen, insbesondere aufgrund ihrer Relevanz für die inhaltliche Ausrichtung der Lehreraus- und -fortbildung.

Des Weiteren ist das faktorenanalytische Verfahren variablenzentriert und berücksichtigt nicht das Mischungsverhältnis bzw. das Auftreten bei bestimmten Personen (Helmke, 2002). Viertens müssen daher differentielle Effekte separat untersucht werden. Interessant wäre, mit dem vorliegenden Datensatz zu ermitteln, inwiefern die Unterrichtswahrnehmungen in Abhängigkeit von der Schulart, der Klassengröße, dem Notendurchschnitt pro Klasse, dem Fach oder der Vorbildung der Lehrkraft variieren. Auch die Eignung der ich- vs. wir-bezogenen Angaben könnte in Abhängigkeit von Moderatorvariablen signifikant variieren, z. B. in Abhängigkeit von der Klassenstufe. Denkbar ist, dass insbesondere jüngeren Schülern Urteile aus der ich-Perspektive schwerer fallen als aus der wir-Perspektive, da sie sich stärker als Ältere an der Meinung der Klasse bzw. Peer-Gruppe orientieren. Die Erkenntnisse aus diesen Untersuchungen könnten zu einer Differenzierung des Instruments in unterschiedliche Versionen führen.

Fünftens verfügt der Datensatz über eine kleine Stichprobe von Lehrkräften, die von mehreren Klassen beurteilt wurden. Dass die gleiche Lehrkraft in unterschiedlichen Klassen unterschiedlich beurteilt werden könnte, ließe sich dadurch erklären, dass das Lehrerverhalten auch eine Folge des Schülerverhaltens und über die Situationen hinweg nicht stabil ist. Inwiefern sich die Angaben unterscheiden und welche Variablen der in der Lehrerstich-

probe erhobenen soziodemographischen Daten, wie z. B. der Lehrberufsdauer oder der bereits in der Klasse unterrichteten Jahre dabei eine Rolle spielen, können Gegenstand einer eigenen Untersuchung sein. Viertens könnte die Lehrerstichprobe hinsichtlich der Eignung der Ergebnismeldung interviewt werden, um dieses wesentliche Instrument im Rahmen des Feedbackprozesses weiter optimieren und den Bedürfnissen der Lehrkräfte anpassen zu können.

Einer *neuen Stichprobe* hingegen bedarf die Überprüfung des Fragebogens im Hinblick auf den Einsatz im Längsschnittdesign. Ditton und Arnold (2004) zeigen, dass bei sehr hoher Stabilität der Schülerurteile über den längsschnittlichen Verlauf, Veränderungen in den Schülereinschätzungen durch das beurteilte Fach und die Schulart, nicht aber durch die Klassenstufe moderiert werden. Diese Befunde könnten in diesem Zuge überprüft und um weitere Variablen hinsichtlich der Unterrichtsentwicklung ergänzt werden. Ebenso muss überprüft werden, inwiefern Verlaufsprofile der Unterrichtsmerkmale, die Veränderungen der Unterrichtswahrnehmungen hinsichtlich der Unterrichtsmerkmale kennzeichnen, in Bezug auf die Veränderungssensibilität des Fragebogens gerechtfertigt sind. Dies bringt wiederum einen Mehrwert für eine systematische Unterrichtsentwicklung.

In Studien können Zusammenhänge zwischen den Schülerperzeptionen und verschiedenen unterrichtlichen Zielkriterien gezeigt werden (De Jong & Westerhof, 2001; Gruehn, 2000). Diesbezüglich wäre m. E. eine Effektzerlegung interessant, d. h. welche Varianzanteile lassen sich kausal auf die hohe Ausprägung des eingeschätzten Unterrichtsmerkmals und welche auf die durch das Evaluationsprozedere bedingte Zuwendung zum Schüler an sich zurückführen. Ist die Effektstärke im zweiten Fall hoch, müssen wichtige Implikationen hinsichtlich der Rolle des Schülers im Evaluationsprozess gezogen werden.

In den kognitiven Pretestinterviews wurde deutlich, dass den Probanden die Diskrimination zwischen einer konkreten Unterrichtsstunde und einem allgemeinen Lehrerurteil schwer fällt. Eine Hypothese dazu wäre, dass Schüler eher über lehrerspezifische Schemata als Resultat einer Aggregation vieler Verhaltensepisoden über die Zeit hinweg verfügen. Clausen (2002) zeigt, dass Schülerurteile einen skalenübergreifenden Varianzanteil beinhalten, der ein globales Beurteilungsschema nahelegt. Auch in der arbeits- und organisationspsychologischen Forschung zeigt sich, dass Beurteiler schnell generalisieren, auch wenn konkret beobachtbares Verhalten erfasst werden soll (Schuler & Höft, 2007). Hier wären weitere Untersuchungen, eventuell auch anhand experimenteller Variationen des



zeitlichen Urteilsbezugs interessant. Kognitive Interviews könnten ebenfalls Aufschluss über die zugrunde liegenden Urteilsprozesse beim Bezug auf eine konkrete, soeben erlebte Unterrichtsstunde vs. den Unterricht im Allgemeinen geben.

Der Fragebogen ist fächerübergreifend konzipiert und erfasst daher keine fachspezifischen Qualitätskomponenten. Eine Erweiterung des Fragebogens um fachspezifische Items in Form eines ‚Baukastentools‘ wäre hilfreich, um die Akzeptanz des Instruments seitens der Lehrkräfte durch eine bessere Passung hinsichtlich fachspezifischer Situationen zu erhöhen. Des Weiteren könnte der Fragebogen bei entsprechenden Items mit Häufigkeitsangaben der zu beurteilenden Sachverhalte ergänzt werden (vgl. Schwartz & Sudman, 1996). Empirische Überprüfungen zum Ausmaß der Übereinstimmung dieser mit externen Beobachtungsdaten und der Selbsteinschätzung der Lehrkraft stehen jedoch noch aus.

## 12.4 Resümee und Ausblick

In erster Linie leistet die Arbeit einen Mehrwert für die schulische Selbstevaluationsarbeit, indem Lehrkräften ein Feedbackinstrument zur Einschätzung der Lehr-Lernprozesse aus der Schülerperspektive zur Verfügung gestellt wird. Der überprüfte Fragebogen erfasst dabei einen Ausschnitt der im Orientierungsrahmen Schulqualität in Rheinland-Pfalz (MBWJK, 2007) kommunizierten prozessbezogenen Unterrichtsqualitätsmerkmale nach Helmke (2007c) mit dem übergeordneten Ziel der Unterrichtsentwicklung. Diesbezüglich erfüllt der Fragebogen eine notwendige, aber keine hinreichende Voraussetzung. Es müssen viele Elemente ineinander greifen, um das Potential des Fragebogens in konkrete Veränderungen in Bezug auf die Lehr-Lernprozesse umsetzen zu können. Insbesondere das bis dato noch nicht im Schulkontext systematisch integrierte Schülerfeedback stellt eine große Herausforderung dar. Der Blick in andere Fachbereiche zeigt, dass die Notwendigkeit des Erhebens der Perspektive der ‚Betroffenen‘ längst erkannt wurde. Im arbeits- und organisationspsychologischen Kontext ist die Mitarbeiterbefragung mittlerweile ein Instrument zeitgemäßer Unternehmensführung und wird zur Ableitung von Hinweisen auf das betriebliche Stärken- und Schwächenprofil als Grundlage konkreter gestalterischer Maßnahmen zur Einleitung von Änderungsprozessen eingesetzt (Bungard & Jöns, 1997). Die Autorin erhofft sich, die in der Regel noch vorhandene Tabureaktion auf das Einholen von *Schüler*-rückmeldungen durch diesen Fragebogen zumindest aufweichen zu können. Ein herausforderndes nächstes Ziel ist in einer treffenden Metapher von MacBeath, Schratz, Meuret und Jakobson umschrieben, dass „Self-evaluation .... should never be viewed as an

imposition, as onerous or time-consuming, but as an essential and integral element of good teaching...” (2000, S. 94). Betrachtet man das Potential, Schüler auch in die Schulentwicklung einzubeziehen, so können als Fernziel die von Burkard, Eikenbusch und Ekholm (2003) beschriebenen Planungs- und Entwicklungsgespräche zwischen Schülern und Lehrkräften gesehen werden, die im schwedischen Schulsystem als ein zentrales Element im Schulentwicklungsmaßnahmenkatalog verankert sind.

Darüber hinaus zeigt die Arbeit inhärent die Notwendigkeit einer engen Vernetzung von Theorie und Praxis im Rahmen der Fragebogenentwicklung auf und appelliert zum verstärkten Einsatz von qualitativen Interviews als Brückenschlag. Auch mit dem vorliegenden Selbstevaluationsverfahren können Theorie und Praxis zusammengeführt werden, indem der auf der Basis einer Aktionsforschung (Altrichter & Posch, 1998) generierte Erkenntnisfortschritt in die Verbesserung der Fragebogenkonstruktion eingeht. Lehrkräfte könnten neben ihren Erfahrungen auch die erhobenen Daten in den wissenschaftlichen Überprüfungsprozess zur Validierung und Optimierung des Fragebogenverfahrens einbringen, was erlaubt, Lehrkräften über einen verbesserten Fragebogen hinaus auch Referenzwerte als eine weitere Feedbackmöglichkeit zur Verfügung zu stellen.

Ferner trägt die durchgeführte Erhebung mit einem umfangreichen Datenpool als Basis-messung im Dreischritt der (1) Standortbestimmung, (2) Intervention und (3) Wiederholungsmessung zur Überprüfung der Interventions- bzw. Maßnahmeneffektivität (Helmke, Piskol, Pikowsky & Wagner, in Druck) zu Folgeuntersuchungen im Rahmen des Projekts Unterrichtsdiagnostik bei. Geplant ist eine Wiederholungsmessung zu einem zweiten Zeitpunkt, um ermitteln zu können, inwiefern die Fragebogenergebnisse über die Handlungsintention hinaus volitionale Planungs- und Handlungsprozesse in Bezug auf die Unterrichtsentwicklung angestoßen haben. Diese Untersuchung würde den Bogen von einer modernen Volitionspsychologie (Heckhausen & Heckhausen, 2006) zur empirisch orientierten Unterrichtsentwicklungsforschung spannen (Helmke et al., 2007).

Aber auch unabhängig von potentiellen zukünftigen Studien kann der in dieser Arbeit erprobte Fragebogen als wegweisend für die Verzahnung empirischer Unterrichtsqualitätsforschung und schulischer Evaluationsarbeit angesehen werden.

### Literaturverzeichnis

- Achtenhagen, F., Sembill, D. & Steinhoff, E. (1979). Die Lehrerpersönlichkeit im Urteil von Schülern. *Zeitschrift für Pädagogik*, 25, 191-208.
- Ahnen, D. (2007). *Neue Schulstruktur in Rheinland-Pfalz fördert Chancengleichheit und Durchlässigkeit, zeigt klare Wege und bietet gute Perspektiven*. Verfügbar unter: <http://hauptschule.bildung-rp.de/neue-schulstruktur.html> [30.10.2007].
- Altrichter, H. (1999). Evaluation als Alltäglichkeit, als Profession und als Interaktion. In J. Thonhauser & R. Albert (Hrsg.), *Evaluation im Bildungsbereich: Wissenschaft und Praxis im Dialog* (S. 103-120). Innsbruck: Studien-Verlag.
- Altrichter, H. & Posch, P. (1998). *Lehrer erforschen ihren Unterricht: Eine Einführung in die Methoden der Aktionsforschung*. Bad Heilbrunn: Klinkhardt.
- Amelang, M. & Schmidt-Atzert, L. (2006). *Psychologische Diagnostik und Intervention*. Heidelberg: Springer.
- Amelang, M. & Zielinski, W. (2002). *Psychologische Diagnostik und Intervention*. Berlin: Springer.
- Anderson, G.J. (1973). *The assessment of learning environments: A manual for the LEI and MCI*. Halifax, N.S.: Atlantic Institute for Education.
- Anderson, G.J. & Walberg, H. J. (1974). Learning environments. In H. J. Walberg (Hrsg.), *Evaluating Educational Performance*. Berkeley: McCutchan.
- Angleitner, A. & Riemann, R. (1996). Selbstberichtsdaten: Fragebogen, Erlebnisanalyse. In K. Pawlik (Hrsg.), *Enzyklopädie der Psychologie, Themenbereich C: Theorie und Forschung, Serie VIII: Differentielle Psychologie und Persönlichkeitsforschung, Band 1: Grundlagen und Methoden der Differentiellen Psychologie*. Göttingen: Hogrefe.
- Arbeitsgruppe Qualitätsstandards (MBWJK Rheinland-Pfalz, IFB). (2007). Qualitätsstandards für Beratungsgruppen für Schulen in Rheinland-Pfalz (Unveröffentlichtes Arbeitspapier). Mainz, Speyer: MBWJK Rheinland-Pfalz, IFB.
- Arbinger, R. & von Saldern, M. (1984). Schulische Umwelt und soziales Klima in Schulklassen. *Psychologie in Erziehung und Unterricht*, 31, 81-99.
- Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides & R. E. Schumacker (Hrsg.), *Advanced structural equation modeling: Issues and techniques* (S. 243-277). Mahwah: Lawrence Erlbaum and Associates.
- Babad, E. (1996). How high is "high inference"? Within classroom differences in students' perceptions of classroom interaction. *Journal of Classroom Interaction*, 31, 1-9.
- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2006). *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung*. Berlin: Springer.
- Balk, M. (2000). *Evaluation von Lehrveranstaltungen - Die Wirkung von Evaluationsrückmeldung*. Frankfurt am Main: Lang.

- Bandura, A. (1976). *Lernen am Modell: Ansätze zu einer sozial-kognitiven Lerntheorie*. Stuttgart: Klett.
- Bastian, J. (2007). *Einführung in die Unterrichtsentwicklung*. Weinheim und Basel: Beltz Verlag.
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9, 469-520.
- Beauducel, A. (2001). Problems with parallel analysis in data sets with oblique simple structure. *Methods of Psychological Research*, 6, 141-157.
- Beauducel, A. & Wittmann, W. W. (2005). Simulation study on fit indices in confirmatory factor analysis based on data with slightly distorted simple structure. *Structural Equation Modeling*, 12, 41-75.
- Berner, H. (1999). *Didaktische Kompetenz*. Bern: Verlag Paul Haupt.
- Bless, H., Fiedler, K. & Strack, F. (2004). *Social cognition: How individuals construct social reality*. Philadelphia: Psychology Press.
- Bless, H. & Schwarz, N. (2002). Konzeptgesteuerte Informationsverarbeitung. In D. Frey & M. Irle (Hrsg.), *Theorien der Sozialpsychologie, Bd. III* (S. 257-278). Bern: Huber.
- Bloom, B. S. (1968). *Learning for mastery. Evaluation Comment 1* (2). Los Angeles: University of California.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley & Sons.
- Borsboom, D. & Mellenbergh, G. J. (2002). True scores, latent variables and constructs: A comment on Schmidt and Hunter. *Intelligence*, 30, 505-514.
- Bortz, J. (1999). *Statistik für Sozialwissenschaftler*. Heidelberg: Springer.
- Brok den, P., Brekelmans, M. & Wubbels, T. (2006). Multilevel issues in research using students' perceptions of learning environments: The case of the Questionnaire on Teacher Interaction. *Learning Environments Research*, 9, 199-213.
- Brophy, J. & Good, T. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Hrsg.), *Handbook of research on teaching* (S. 328-375). New York: Macmillan.
- Brosius, F. (2002). *SPSS 11*. Bonn: mitp Verlag.
- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Studium.
- Buhren, C.-G., Killus, D. & Müller, S. (2000). Implementation und Wirkung von Selbstevaluation in Schulen. In H.-G. Rolff, W. Bos, K. Klemm, H. Pfeiffer & R. Schulz-Zander (Hrsg.), *Jahrbuch der Schulentwicklung, Bd. 11* (S. 327-364). Weinheim: Juventa.
- Bungard, W. & Jöns, I. (1997). *Mitarbeiterbefragungen. Ein Instrument des Qualitäts- und Innovationsmanagements*. Weinheim: Psychologie Verlags Union.
- Bürgisser, D. (2007). *Warum nicht Unterrichtsbeurteilungen durch Schüler?* Verfügbar unter: <http://www.mvz.ch/qi/pdf/QiLQS.pdf> [10.11.2007].

- Burkard, C., Eikenbusch, G. & Ekholm, M. (2003). *Starke Schüler - gute Schulen. Wege zu einer neuen Arbeitskultur im Unterricht*. Berlin: Cornelsen.
- Burkhard, C. (1997). Externe Evaluation: Rückenwind oder Motivationskiller. *Pädagogik*, 5, 10-13.
- Burkhard, C. & Eikenbusch, G. (2000). *Praxishandbuch Evaluation in der Schule*. Berlin: Cornelsen Scriptor.
- Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum.
- Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Carroll, J. B. (1964). A model of school learning. *Teachers College Record*, 64, 723-733.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 245-276.
- Cattell, R. B., Eber, H. W. & Tatsuoka, M. M. (1970). *Handbook for the sixteen Personality Factor Questionnaire (16 PF)*. Champaign, IL: Institute for Personality and Ability Testing.
- Centra, J. A. (1993). *Reflective faculty evaluation: Enhancing teaching and determining faculty effectiveness*. San Francisco: Jossey-Bass.
- Clausen, M. (2002). *Unterrichtsqualität - Eine Frage der Perspektive?* Münster: Waxmann.
- Clausen, M., Reusser, K. & Klieme, E. (2003). Unterrichtsqualität auf der Basis hoch-inferenter Unterrichtsbeurteilungen: Ein instruktionspsychologischer Vergleich zwischen Deutschland und der deutschsprachigen Schweiz. *Unterrichtswissenschaft*, 31, 122-141.
- Clausen, M., Schnabel, K. & Schröder, K. (2002). Konstrukte der Unterrichtsqualität im Expertenurteil. *Unterrichtswissenschaft*, 30, 246-260.
- Cohen, J. (1992). A Power Primer. *Psychological Bulletin*, 112, 155-159.
- Cohen, J., Cohen, P., West, S. G. & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah: Lawrence Erlbaum Associates.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D. & York, R. L. (1966). *Equality of educational opportunity*. Washington: National Center for Educational Statistics.
- Conrad, W. (2003). Grundlagen der Klassischen Messtheorie, Unveröffentlichtes Begleitskriptum zur Vorlesung: Universität Mannheim.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78, 98-104.
- Creemers, B. P. M. (1994). Effective instruction: An empirical basis for a theory of educational effectiveness. In D. Reynolds (Hrsg.), *Advances in school effectiveness research and practice* (S. 189-205). Oxford: Pergamon.

- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Cronbach, L. J., Ambron, S. R., Dornbusch, S. M., Hess, R. D., Hornik, R. C., Phillips, D. C., Walker, D. F. & Weinert, S. S. (1980). *Toward reform of program evaluation*. San Francisco: Jossey-Bass.
- Daumen, C., Miethner, J. & Pikowsky, B. (2005). *Qualitätsrahmen für Schulen in Rheinland Pfalz*. Internes Arbeitspapier des Ministeriums für Bildung, Wissenschaft, Jugend und Kultur, Rheinland-Pfalz.
- De Jong, R. & Westerhof, K. J. (2001). The quality of student ratings of teacher behaviour. *Learning Environments Research*, 4, 51-85.
- Deci, E. L. & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum Press.
- Ditton, H. (1998). *Mehrebenenanalyse*. Weinheim: Juventa.
- Ditton, H. (2000). Qualitätskontrolle und -sicherung in Schule und Unterricht - ein Überblick zum Stand der empirischen Forschung. In A. Helmke, W. Hornstein & E. Terhart (Hrsg.), *Qualitätssicherung im Bildungsbereich. Beiheft Nr. 41 der Zeitschrift für Pädagogik* (S. 73-92). Weinheim: Beltz.
- Ditton, H. (2002a). Lehrkräfte und Unterricht aus Schülersicht: Ergebnisse einer Untersuchung im Fach Mathematik. *Zeitschrift für Pädagogik*, 48, 262-286.
- Ditton, H. (2002b). Unterrichtsqualität – Konzeptionen methodische Überlegungen und Perspektiven. *Unterrichtswissenschaft*, 30, 197-212.
- Ditton, H. & Arnoldt, B. (2004). Wirksamkeit von Schülerfeedback zum Fachunterricht. In M. Prenzel & J. Doll (Hrsg.), *Bildungsqualität von Schule: Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategien der Qualitätsverbesserung* (S. 152-173). Münster: Waxmann.
- Ditton, H. & Kreckler, L. (1995). Qualität von Schule und Unterricht: Empirische Befunde zu Fragestellungen und Aufgaben der Forschung. *Zeitschrift für Pädagogik*, 41, 507-529.
- Dorenkamp, L. (1999). Evaluation als Steuerungs- und Führungsinstrument für (berufliche) Schulen. In L. Dorenkamp & K. Lauks (Hrsg.), *Schulentwicklung durch Bildungsinnovationen: Ergebnisse eines internationalen Symposiums zum Modellversuch DOPKAU* (S. 181-190). Rostock: Universität Rostock, Lehrstuhl für Wirtschaftspädagogik.
- Dreesmann, H. (1978). *Unterrichtsklima als situative Bedingung für kognitive Prozesse und das Leistungsverhalten von Schülern*. Dissertation an der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität Heidelberg.
- Dreesmann, H. (1979). *Unterrichtsklima als situative Bedingung für kognitive Prozesse und das Leistungsverhalten von Schülern*. Dissertation an der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität Heidelberg.
- Dreesmann, H. (1982). *Unterrichtsklima: Wie Schüler den Unterricht wahrnehmen*. Weinheim: Beltz.

- Eder, F. (1992). Schulklima und Entwicklung allgemeiner Interessen in der Schule. In A. Krapp & M. Prenzel (Hrsg.), *Interesse, Lernen, Leistung* (S. 165-194). Münster: Aschendorff.
- Eder, F. (1996). *Schul- und Klassenklima*. Wien: Studien-Verlag.
- Eder, F. (2001). Schul- und Klassenklima. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 578-586). Weinheim: Beltz.
- Eder, F. (2002). Unterrichtsklima und Unterrichtsqualität. *Unterrichtswissenschaft*, 30, 213-228.
- Einsiedler, W. (1997). Unterrichtsqualität und Leistungsentwicklung: Literaturüberblick. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 226-240). Weinheim: Psychologie Verlags Union.
- Einsiedler, W. (2002). Das Konzept "Unterrichtsqualität". *Unterrichtswissenschaft*, 30, 194-196.
- Elsner, G. & Börner, F. (2000). *Unterrichtsbeurteilung durch Schüler - ein Pilotprojekt im Auftrag des Sächsischen Staatsministeriums*. Dresden: Sächsische Arbeitsstelle für Schule und Jugendhilfe.
- Fend, H. (1977). *Schulklima*. Weinheim: Beltz.
- Fend, H. (1981). *Theorie der Schule*. München: Urban & Schwarzenberg.
- Fischer, G. H. (1974). *Einführung in die Theorie psychologischer Tests: Grundlagen und Anwendungen*. Bern: Hogrefe.
- Fisseni, H. J. (1997). *Lehrbuch der psychologischen Diagnostik*. Göttingen: Hogrefe.
- Follman, J., Lucoff, M., Small, L. & Power, F. (1974). Kinds of keys of student ratings of faculty teaching effectiveness. *Research in Higher Education*, 2, 173-179.
- Fraser, B. J., Anderson, G. J. & Walberg, H. J. (1982). *Assessment of learning environments: Manual for learning environment inventory (LEI) and my class inventory (MCI)*. Bentley: Western Australien Institut of Technology.
- Fraser, B. J. & Walberg, H. J. (1981). Psychosocial learning environment in science classrooms: A review of research. *Studies in Science Evaluation*, 8, 67-92.
- Fraser, B. J., Walberg, H. J., Welch, W. W. & Hattie, J. A. (1987). Syntheses of educational productivity research. *International Journal of Educational Research*, 11, 147-252.
- Getzels, J. W. & Thelen, H. (1960). The classroom group as a unique social system. In N. B. Henry (Hrsg.), *The dynamics of instructional groups: Sociopsychological aspects of teaching and learning*. Chicago: University of Chicago Press.
- Greene, J. C. (1988). Stakeholder participation and utilization in program evaluation. *Evaluation Review*, 12, 91-116.
- Gruehn, S. (1995). Vereinbarkeit kognitiver und nichtkognitiver Ziele im Unterricht. *Zeitschrift für Pädagogik*, 41, 531-553.
- Gruehn, S. (2000). *Unterricht und schulisches Lernen*. Münster, München, Berlin [u.a.]: Waxmann.

- Harnischfeger, A. & Wiley, D. E. (1977). Kernkonzepte des Schullernens. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 9, 207-228.
- Hartley, S. S. (1977). Meta-analysis of the effects of individually paced instruction in mathematics. *Dissertation Abstracts International*, 38, 4003A, University Microfilms No. 77-29, 926.
- Heckhausen, J. & Heckhausen, H. (2006). *Motivation und Handeln*. Heidelberg: Springer.
- Helmke, A. (2002). Kommentar: Unterrichtsqualität und Unterrichtsklima: Perspektiven und Sackgassen. *Unterrichtswissenschaft*, 30, 261-277.
- Helmke, A. (2003). *Unterrichtsqualität erfassen, bewerten, verbessern*. Seelze: Kallmeyer.
- Helmke, A. (2006a). *Unterrichtsqualität: Erfassen, Bewerten, Verbessern*. Seelze: Kallmeyersche Verlagsbuchhandlung.
- Helmke, A. (2006b). Was wissen wir über guten Unterricht? Über die Notwendigkeit einer Rückbesinnung auf den Unterricht als dem „Kerngeschäft“ der Schule. *Pädagogik*, 2, 42-45.
- Helmke, A. (2007a). Einblick in die Lehr-Lern-Situation. In T. Riecke-Baulecke (Hrsg.), *SchulleitungPlus: Schule und Unterricht erfolgreich gestalten* (S. 88-101/208-209). München: Oldenbourg.
- Helmke, A. (2007b). *Unterrichtsbeobachtungsbogen (UBB), Version 3.0*. Landau: Universität Koblenz-Landau.
- Helmke, A. (2007c). *Was wissen wir über guten Unterricht?* Verfügbar unter: <http://www.iqesonline.net/index.cfm?id=94923064-e0c6-b4e6-2d2f-1cc64ad1d9d8> [15.09.2007].
- Helmke, A. & AQS, Wissenschaftliche Beirat (2007). *Unterrichtsbeobachtungsbogen Einblicknahmen in Lehr- und Lernsituationen (ELL V3.0)*. Mainz: AQS.
- Helmke, A., Helmke, T., Heyne, N., Hosenfeld, A., Hosenfeld, I., Schrader, F.-W. & Wagner, W. (2008). Zeitnutzung im Grundschulunterricht: Ergebnisse der Unterrichtsstudie "VERA - Gute Unterrichtspraxis". *Zeitschrift für Grundschulforschung*, 1, 23-36.
- Helmke, A., Helmke, T., Schrader, F.-W., Wagner, W., Kleinbub, I. & Pikowsky, B. (2007). *Projektantrag: Diagnostik des Unterrichts als Bestandteil diagnostischer Kompetenz und pädagogischer Professionalität*. Unveröffentlichter Projektantrag, Universität Koblenz-Landau in Kooperation mit der Graduiertenschule der Exzellenz 'Unterrichtsprozesse' und dem Institut für schulische Fortbildung und schulpсихologische Beratungsstelle, Speyer.
- Helmke, A. & Hosenfeld, I. (2005). Standardbezogene Unterrichtsevaluation. In G. Brägger, B. Bucher & N. Landwehr (Hrsg.), *Schlüsselfragen zur externen Schulevaluation* (S. 127-151). Bern: h.e.p.-Verlag.
- Helmke, A., Piskol, K., Pikowsky, B. & Wagner, W. (in Druck). Schüler als Experten für Unterricht: Entwicklung eines Schülerfragebogens zur Unterrichtsqualität. *Lernende Schule*.
- Helmke, A., Schneider, W. & Weinert, F. E. (1986). Quality of instruction and classroom learning outcomes: The German contribution to the IEA Classroom Environment Study. *Teaching and Teacher Education*, 2, 1-18.



- Helmke, A. & Schrader, F.-W. (1993). Was macht erfolgreichen Unterricht aus? Ergebnisse der Münchner Studie. *Praxis Schule, 1*, 11-13.
- Helmke, A. & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistungen. In F. E. Weinert (Hrsg.), *Enzyklopädie der Psychologie, Band 3 (Psychologie der Schule und des Unterrichts)* (S. 71-176). Göttingen: Hogrefe-Verlag.
- Helmke, T., Helmke, A., Schrader, F.-W., Wagner, W., Nold, G. & Schröder, K. (2008). Die Videostudie des Englischunterrichts. In DESI-Konsortium (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 345-363). Weinheim: Beltz.
- Hense, J. U. (2006). *Selbstevaluation: Erfolgsfaktoren und Wirkungen eines Ansatzes zur selbstbestimmten Qualitätsentwicklung im schulischen Bereich*. Frankfurt am Main: Lang.
- Hofer, M. (1981). Lehrerverhalten aus der Sicht des Schülers. *Pädagogische Welt, 35*, 49-56.
- Hofer, M. (1982). Lehrerverhalten aus der Sicht des Schülers. *Unterrichtswissenschaft, 10*, 240-251.
- Horster, L. & Rolff, H.-G. (2001). *Unterrichtsentwicklung: Grundlagen, Praxis, Steuerungsprozesse*. Weinheim: Beltz.
- Jencks, C., Smith, M., Acland, H., Bane, M. J., Cohen, D., Gintis, H., Heyns, B. & Michelson, S. (1972). *Inequality: A reassessment of the effect of family and schooling in America*. New York: Basic Books.
- Kahl, T. N. (1977). *Unterrichtsforschung. Probleme, Methoden und Ergebnisse der empirischen Untersuchung unterrichtlicher Lernsituationen*. Kronberg/Ts.: Scriptor-Verlag.
- Kahl, T. N., Buchmann, M. & Witte, E. H. (1977). Ein Fragebogen zur Schülerwahrnehmung unterrichtlicher Lernsituationen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 9*, 277-285.
- Kemme, S. (1998). *Lehrer- und Schülerwahrnehmung von Unterricht und Unterrichtsklima. Ein Perspektivenvergleich. Unveröffentlichte Diplomarbeit*. Freie Universität Berlin.
- Kempfert, G. & Rolff, H.-G. (2005). *Qualität und Evaluation. Ein Leitfaden für Pädagogisches Qualitätsmanagement*. Weinheim: Beltz.
- Kenny, D. A. (2004). PERSON: A general model of interpersonal perception. *Personality and Social Psychology Review, 8*, 265-280.
- Kersting, M. (2003). Augenscheinvalidität. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der Psychologischen Diagnostik* (S. 54-55). Weinheim: Beltz.
- Klieme, E. & Rakoczy, K. (2003). Unterrichtsqualität aus Schülerperspektive: Kulturspezifische Profile, regionale Unterschiede und Zusammenhänge mit Effekten von Unterricht. In Deutsches PISA-Konsortium (Hrsg.), *PISA 2000 - Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (S. 333-360). Opladen: Leske + Budrich.
- König, J. (2000). *Einführung in die Selbstevaluation*. Freiburg: Lambertus.
- Krapp, A. & Weidenmann, B. (2001). *Pädagogische Psychologie*. Weinheim: Beltz PVU.

- Kromrey, H. (2001). Evaluation – ein vielschichtiges Konzept. Begriff und Methodik von Evaluierung und Evaluationsforschung. Empfehlungen für die Praxis. *Sozialwissenschaften und Berufspraxis*, 24, 105-131.
- Krug, S. & Lecybyl, R. (1999). Die Wirkung experimentell variierten Lehrerverhaltens auf Unterrichtswahrnehmung, Lernbereitschaft und Leistung von Schülern. In F. Rheinberg & S. Krug (Hrsg.), *Motivationsförderung im Schulalltag* (S. 57-73). Göttingen: Hogrefe.
- Kunter, M. & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research*, 9, 231-251.
- Kurz, K., Prüfer, P. & Rexroth, M. (1999). Zur Validität von Fragen in standardisierten Erhebungen. Ergebnisse des Einsatzes eines kognitiven Pretestinterviews. *ZUMA-Nachrichten*, 44, 83-107.
- Ladwig, D. H. & Domsch, M. E. (1997). Mitarbeiterbefragungen als Instrument eines 'Culture Change' im Unternehmen. In W. Bungard & I. Jöns (Hrsg.), *Mitarbeiterbefragung. Ein Instrument des Innovations- und Qualitätsmanagements* (S. 74-83). Weinheim: Beltz/PVU.
- Lance, C. E., LaPointe, J. A. & Stewart, H. M. (1994). A test of the context dependency of three causal models of halo rater error. *Journal of Applied Psychology*, 79, 332-340.
- Lange, B., Kuffner, H. & Schwarzer, R. (1983). *Schulangst und Schulverdrossenheit. Eine Längsschnittanalyse von schulischen Sozialisationseffekten*. Opladen: Westdeutsche Verlagsgesellschaft.
- Lange, H. (2001). Die bildungspolitische Bedeutung von Schulleistungsvergleichen. In G. Kaiser, N. Knoche, D. Lind & W. Zillmer (Hrsg.), *Leistungsvergleiche im Mathematikunterricht* (S. 1-28). Hildesheim: Franzbecker.
- Lewin, K. (1936). *Feldtheorie in den Sozialwissenschaften*. Bern: Huber.
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse*. Weinheim: Beltz.
- Luyten, H. (2003). The size of school effects compared to teacher effects: An overview of the research literature. *School Effectiveness and School Improvement*, 14, 31-51.
- MacBeath, J., Schratz, M., Meuret, D. & Jakobsen, L. (2000). *Self-evaluation in European schools: A story of change*. London: Routledge.
- MacCallum, R. C., Widman, K. F., Zhang, S. & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4, 87-99.
- Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57, 519-530.
- Marsh, H.W. (1984). Student evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, 76, 707-754.
- MBWJK, Rheinland-Pfalz. (2007). *Orientierungsrahmen Schulqualität für Rheinland-Pfalz*. Mainz: Ministerium für Bildung, Wissenschaft, Jugend und Kultur Rheinland-Pfalz.
- Mendoza, J. L., Stafford, K. L. & Stauffer, J. M. (2000). Large-sample confidence intervals for validity and reliability coefficients. *Psychological Methods*, 5, 356-369.

- Michel, L. & Conrad, W. (1982). Testtheoretische Grundlagen psychometrischer Tests. In K. J. Groffmann & L. Michel (Hrsg.), *Enzyklopädie der Psychologie, Themenbereich B: Methodologie und Methoden, Serie II: Psychologische Diagnostik, Band 1: Grundlagen Psychologischer Diagnostik* (S. 1-129). Göttingen: Hogrefe.
- Moos, R. H. (1979). *Evaluating educational environments*. San Francisco, CA: Jossey-Bass.
- Moos, R. H. & Trickett, E. J. (1974). *Classroom environment scale manual*. Palo Alto: Consulting Psychologists Press.
- Moosbrugger, H. & Kelava, A. (2007). *Testtheorie und Fragebogenkonstruktion*. Heidelberg: Springer.
- Müller-Kohlenberg, H. (2004). *Empfehlungen zur Anwendung der Standards für Evaluation im Handlungsfeld der Selbstevaluation (DeGEval)*. Verfügbar unter: <http://www.degeval.de/calimero/tools/proxy.php?id=172> [22.08.2008].
- Mummendey, H. D. (1995). *Die Fragebogen-Methode*. Göttingen: Hogrefe.
- Murphy, K. R. & Davidshofer, C. O. (2001). *Psychological testing: Principles and applications*. Upper Saddle River, NJ: Prentice Hall.
- Murray, H. A. (1938). *Explorations in personality*. New York: Oxford University Press.
- Muthén, B. O. & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology*, 45, 265-277.
- Niegemann, H. (2001). Lehr-Lern-Forschung. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 387-393). Weinheim: Beltz.
- O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments & Computers*, 32, 396-402.
- O'Sullivan, R. G. & D'Agostino, A. (2002). Promoting evaluation through collaboration. *Evaluation*, 8, 372-387.
- OECD, Organization for Economic Cooperation and Development. (1989). *Schools and quality. An international report*. Paris: OECD.
- Pekrun, R. (1985). Schulklima. In W. Twellmann (Hrsg.), *Handbuch Schule und Unterricht* (Band 7.1, S. 524-547). Düsseldorf: Schwann.
- Pikowsky, B. (2007a). Qualifizierung der Evaluationsteams AQS (Unveröffentlichtes Manuskript). Speyer: IFB.
- Pikowsky, B. (2007b). Wer wird über was orientiert? Qualitätsrahmen und Orientierungsrahmen schulischer Qualität. *Lernende Schule*, 40, 9-13.
- Preston, C. C. & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104, 1-15.

- Rindermann, H. (2001). Die studentische Beurteilung von Lehrveranstaltungen - Forschungsstand und Implikationen. In C. Spiel (Hrsg.), *Evaluation universitärer Lehrer - zwischen Qualitätsmanagement und Selbstzweck* (S. 61-88). Münster: Waxmann.
- Rosenstiel, L. von. (1999). Die "lernende Organisation" als Ausgangspunkt für Qualitätsentwicklung. In F. Peterander & O. Speck (Hrsg.), *Qualitätsmanagement in sozialen Einrichtungen* (S. 41-62). München: Ernst Reinhardt.
- Rost, D. (1996). *Lehrbuch Testtheorie, Testkonstruktion*. Göttingen: Hogrefe.
- Rutter, M., Maughan, B., Mortimore, P. & Ouston, J. (1979). *Fifteen thousand hours. Secondary schools and their effects on children*. Cambridge: Harvard University Press.
- Saldern, M. von & Littig, K.-E. (1987). *Landauer Skalen zum Sozialklima: 4. - 13. Klassen, LASSO 4-13*. Weinheim: Beltz.
- Scheerens, J. (1992). *Effective Schooling: Research, theory and practice*. London: Cassell.
- Schermelleh-Engel, K., Moosbrugger, H. & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8, 23-74.
- Schmid, J. & Leiman, J. N. (1957). The development of hierarchical factor solution. *Psychometrika*, 22, 53-61.
- Schnabel, K. (2001). Psychologie der Lernumwelt. In A. Krapp & B. Weidenmann (Hrsg.), *Pädagogische Psychologie* (S. 467-511). Weinheim: Beltz PVU.
- Schneewind, J. (2007). *Wie Lehrkräfte mit Ergebnissrückmeldungen aus Schulleistungsstudien umgehen*. Dissertation, FB Erziehungswissenschaften und Psychologie, Freie Universität Berlin.
- Schratz, M. (1999). Erreichtes sichern - Neues entwickeln. Selbstevaluation als Bemühen, Qualität zu verstehen. In U. Steffens & T. Bargel (Hrsg.), *Qualitätsentwicklung und Qualitätssicherung von Schulen - Strategien, Instrumente und Erfahrungen* (S. 63-74). Wiesbaden: Hessisches Landesinstitut für Pädagogik.
- Schröder, S. (1999). *Konstrukte der Unterrichtsqualität im Expertenurteil. Unveröffentlichte Diplomarbeit* Freie Universität Berlin.
- Schuler, H. & Höft, S. (2007). Diagnose beruflicher Eignung und Leistung. In H. Schuler (Hrsg.), *Lehrbuch Organisationspsychologie* (S. 289-343). Bern: Huber.
- Schwartz, N. & Sudman, S. (1996). *Answering questions: Methodology for determining cognitive and communicative processes in survey research*. San Francisco, CA: Jossey-Bass.
- Scriven, M. (1991). *Evaluation thesaurus*. Newbury Park: Sage.
- Slavin, R. E. (1994). Quality, appropriateness, incentive, and time: A model of instructional effectiveness. *International Journal of Educational Research*, 21, 141-157.
- Snow, R. E. & Swanson, J. (1992). Instructional psychology, aptitude, adaptation, and assessment. *Annual Review of Psychology*, 43, 583-626.

- Staatskanzlei, Rheinland-Pfalz. (2004). *Schulgesetz (SchulG) vom 30. März 2004*. Verfügbar unter: [http://rlp.juris.de/rlp/SchulG\\_RP\\_2004\\_rahmen.htm](http://rlp.juris.de/rlp/SchulG_RP_2004_rahmen.htm) [01.10.2007].
- Steltmann, K. (1992). Lehrerbeurteilung und Lehrereffektivität. In K. Ingenkamp, R. S. Jäger, H. Petillon & B. Wolf (Hrsg.), *Empirische Pädagogik, Band II*. Weinheim: Deutscher Studienverlag.
- Stern, G. G., Stein, M. A. & Bloom, B. S. (1956). *Methods in personality assessment*. Glencoe, IL: Free Press.
- Stringfield, S. (1994). A model of elementary school effects. In D. Reynolds, B. P. M. Creemers, P. S. Nesselrodt, E. C. Schaffer, S. Stringfield & C. Teddlie (Hrsg.), *Advances in school effectiveness research and practice* (S. 153-187). Oxford: Pergamon.
- Sudman, S., Bradburn, N. M. & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco, Calif.: Jossey-Bass.
- Taylor, S. E. (1981). The interface of cognitive and social psychology. In J. Harvey (Hrsg.), *Cognition, social behavior, and the environment* (S. 189-211). Hillsdale, NJ: Erlbaum.
- Teddlie, C. & Reynolds, D. (2001). *The international handbook of school effectiveness research*. London: Routledge/Falmer.
- Tent, L. & Stelzl, I. (1993). *Pädagogisch-psychologische Diagnostik, Band 1: Theoretische und methodische Grundlagen*. Göttingen: Hogrefe.
- Terhart, E. (2000). Qualität und Qualitätssicherung im Schulsystem. *Zeitschrift für Pädagogik*, 46, 809-829.
- Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, 4, 25-29.
- Treiber, B. (1982). Lehr- und Lernzeiten im Unterricht. In B. Treiber & F. E. Weinert (Hrsg.), *Lehr-Lern-Forschung* (S. 12-36). München: Urban und Schwarzenberg.
- Wagner, W. (2008). *Methodenprobleme bei der Analyse der Unterrichtswahrnehmung aus Schülersicht – am Beispiel der Studie DESI (Deutsch Englisch Schülerleistungen International) der Kultusministerkonferenz*. Dissertation am Fachbereich 8, Psychologie, Universität Koblenz-Landau.
- Waibel, R. (2003). Wissenschaftlich fundierte Instrumente für die Beurteilung der Unterrichtsqualität. *Schulleitung und Schulentwicklung*, 2.
- Walberg, H. J. (1966). *Classroom Climate Questionnaire*. Cambridge, MA: Harvard University.
- Walberg, H. J. (1981). A psychological theory of educational productivity. In F. H. Farley & N. J. Gordon (Hrsg.), *Psychology and education. The state of union* (S. 81-108). Berkely, CA: McCutchan.
- Walberg, H. J. (1990). A theory of educational productivity: Fundamental substance and method. In P. Vedder (Hrsg.), *Fundamental studies in educational research* (S. 19-34). Lisse: Swets & Zeitlinger.

- Walberg, H. J. & Haertel, G. D. (1980). Validity and use of educational environment assessments. *Studies in Educational Evaluation*, 6, 225-238.
- Wang, M. C., Haertel, G. D. & Walberg, H. J. (1990). What influences learning? A content analysis of review literature. *Journal of Educational Research*, 84, 30-43.
- Wang, M. C., Haertel, G. D. & Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research*, 63, 249-294.
- Watkins, M. W. (2004). *Monte Carlo PCA for Parallel Analysis (Software)*. Verfügbar unter: <http://www.public.asu.edu/~mwwatkin/Watkins3.html> [15.05.2008].
- Weinert, F. E., Schrader, F.-W. & Helmke, A. (1989). Quality of instruction and achievement outcomes. *International Journal of Educational Research*, 13, 895-914.
- Weinert, F. E. & Treiber, B. (1982). Einleitung. In B. Treiber & F. E. Weinert (Hrsg.), *Lehr-Lern-Forschung*. München: Urban & Schwarzenberg.
- Wesseler, M. (1999). Evaluation und Evaluationsforschung. In R. Tippelt (Hrsg.), *Handbuch Erwachsenenbildung/Weiterbildung* (S. 737-752). Opladen: Leske + Budrich.
- West, S. G., Finch, J. F. & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Hrsg.), *Structural equation modeling: Concepts, issues, and applications* (S. 56-75). Thousand Oaks, CA: Sage Publications.
- Westermann, R. (2002). Merkmale und Varianten von Evaluationen: Überblick und Klassifikation. *Zeitschrift für Psychologie*, 210, 4-26.
- Widmer, T. (2000). Qualität der Evaluation - Wenn Wissenschaft zur praktischen Kunst wird. In R. Stockmann (Hrsg.), *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder* (S. 77-102). Opladen: Leske + Budrich.
- Will, H., Winteler, A. & Krapp, A. (1987). Von der Erfolgskontrolle zur Evaluation. In H. Will, A. Winteler & A. Krapp (Hrsg.), *Evaluation in der beruflichen Aus- und Weiterbildung*. Heidelberg: Sauer.
- Wittmann, W. W. (1990). Brunswik-Symmetrie und die Konzeption der Fünf-Datenboxen. Ein Rahmenkonzept für umfassende Evaluationsforschung. *Zeitschrift für Pädagogische Psychologie*, 4, 241-251.
- Wottawa, H. J. & Thierau, H. (2003). *Lehrbuch Evaluation*. Bern: Huber.
- Wulf, C. (1975a). Funktionen und Paradigmen der Evaluation. In K. Frey (Hrsg.), *Curriculum-Handbuch, Band 2* (S. 580-600). München: Piper.
- Wulf, C. (1975b). Planung und Durchführung von Evaluation von Curricula und Unterricht. In K. Frey (Hrsg.), *Curriculum-Handbuch, Band 2* (S. 567-580). München: Piper.
- Zwick, W. R. & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 99, 432-442.

## Anhang A Fragebogenimplementierung

### A.1 Kurzbeschreibung zum Fragebogen

Tabelle A.1-1: *Kurzbeschreibung zum Fragebogen*

<b>Kurzbeschreibung ,Schülerfeedbackbogen zur Unterrichtsrückmeldung'</b>	
ZIELGRUPPE	Feedbackinstrument für Lehrkräfte aller Schularten der Sekundarstufe I als Hilfsmittel zur Professionalisierung der Unterrichtstätigkeit
ANWENDUNGS- BEREICH	Erfassen der Schülereinschätzungen zu ausgewählten Merkmalen der Unterrichtsqualität
ERHEBUNGS- METHODE	schriftlich per Fragebogen mit 4-stufigem Antwortformat (stimme zu bis stimme nicht zu)
FRAGEBOGEN- EIGENSCHAFTEN	fächerübergreifende, verhaltensbasierte und i.d.R. ich-bezogen formulierte Aussagen zur Einschätzung einer konkreten Unterrichtsstunde
VERSIONEN	Schülerfragebogen liegt in Version mit weiblicher und männlicher Lehrkraftbezeichnung sowie als Lehrerfragebogen vor
DURCH- FÜHRUNG	zu Beginn der zweiten Doppelstundenhälfte oder der Folgestunde in Absprache mit der Kollegin/dem Kollegen
UMFANG	Fünf Merkmalsbereiche mit insgesamt 60 Aussagen mit einer Bearbeitungsdauer von ca. 15 Minuten (inkl. Instruktion)
ANWENDUNGS- PRINZIPIEN	regelmäßiger Einsatz des Fragebogens, wobei Beteiligung der Schüler/innen freiwillig und anonym ist
TIPP	Einsatz des Lehrerfragebogens mit inhaltsgleichen Items zur Erfassung der Selbsteinschätzung und Abgleich mit der Schülerperspektive
AUSWERTUNG	Excel-Auswertungsmaske zur Eingabe der Fragebogendaten liefert eine Profildarstellung der Ergebnisse auf Klassenebene
ERGEBNIS- RÜCKMELDUNG	Rückmeldung an Schüler/innen im gemeinsamen Dialog, optional Austausch mit (Fach-)Kolleginnen/Kollegen
FRAGEBOGEN- PROJEKTTEAM	Prof. Dr. Andreas Helmke, Universität Koblenz-Landau, Dr. Birgit Pikowsky, IFB, Speyer, Dr. Wolfgang Wagner, Universität Koblenz-Landau und Kathleen Piskol, Universität Mannheim

## A.2 Kommentierte Literaturliste zum Schülerfeedbackeinsatz

**Burkhard, C., Eikenbusch, G. & Ekholm, M. (2003). *Starke Schüler – gute Schulen. Wege zu einer neuen Arbeitskultur im Unterricht*. Berlin: Cornelsen.**

Welche Argumente es für eine Schulentwicklung mit Schülern gibt, wie Schülerrückmeldungen als Kern der Schulentwicklung gestaltet werden können und die Feedbackarbeit mit Schülern gelingt, zeigen die Autoren in übersichtlicher und verständlicher Art und Weise in Kombinationen mit vielfältigen praktischen Tipps und Strategien auf. Dieses Buch ist sowohl der Lehrkraft als auch der Schulleitung ein hilfreicher Ratgeber.

**Bastian, J., Combe, A. & Langer, A. (2007). *Feedbackmethoden. Erprobte Konzepte, evaluierte Erfahrungen*. Weinheim: Beltz.**

Die Autoren stellen ausgehend von einer Einführung zur Schülerrückmeldung als Beitrag zur Unterrichtsentwicklung eine Reihe von Methoden für die Feedbackarbeit im Schulalltag vor. Dabei zeigen sie sowohl verschiedene Varianten der Durchführung der Methoden als auch ihre Vorteile und weitere zu beachtende Aspekte in einer sehr strukturierten und übersichtlichen Form auf.

**Burkhard, C. & Eikenbusch, G. (2000). *Praxishandbuch Evaluation in der Schule*. Berlin: Cornelsen Scriptor.**

Dieses Handbuch beleuchtet die schulische (Selbst-)Evaluation als Ausgangspunkt zur Weiterentwicklung des Unterrichts bzw. der eigenen pädagogischen Tätigkeit. Hervorzuheben ist die Praxisnähe des Buchs, denn es entstand im Rahmen der Durchführung der nordrhein-westfälischen Fortbildungsmaßnahme „Schulentwicklung und Schulaufsicht – Qualitätsentwicklung und -sicherung von Schule“ (QUESS).

**Pikowsky, B. (2004). *Werkstattheft - Schulische Qualitätsarbeit: Anregungen und Beispiele zur Evaluation (Reihe: Schulisches Qualitätsmanagement)*. Speyer: Institut für schulische Fortbildung und schulpsychologische Beratung des Landes Rheinland-Pfalz (IFB).**

Die Autorin stellt wesentliche Evaluationsaspekte im Rahmen schulischer Qualitätsarbeit praxisorientiert als Unterstützungsangebot für Schulen zusammen. Der Erklärung grundlegender (Evaluations-)Konzepte wird dieses Heft genauso gerecht wie der Darstellung konkreter praktischer Instrumente zur Datenerhebung und -auswertung.

**Radnitzky, E. & Schratz, M. (1999). *Der Blick in den Spiegel: Texte zur Praxis von Selbstevaluation und Schulentwicklung*. Innsbruck: StudienVerlag.**

In diesem Buch bearbeiten verschiedene Autoren unterschiedliche Aspekte der Qualitätsentwicklung durch Selbstevaluation. Burkard geht in seinem Kapitel (S. 203-228) von Fallszenen interner Evaluationen aus und leitet daraus interpretativ Dos and Dont's der Selbstevaluation ab.

**Riecke-Baulecke, T. (2004). *Schule Plus. Managementmodell für wirksame Qualitätsentwicklung*. München: Oldenbourg.**

In seinem Buch beschreibt Riecke-Baulecke unter dem Rückgriff auf seine langjährigen Erfahrungen aus der wissenschaftlichen Begleitung von Schulentwicklungsprogrammen Grundlagen und zahlreiche Techniken sowie Strategien u. a. zur Verbesserung der Unterrichtswirksamkeit im Rahmen einer praktischen Schulentwicklungsarbeit.



## Anhang B Qualitative Voruntersuchung

### B.1 SFB-Vorversion

Fassung vom 18.10.07

# Fragebogen für Schülerinnen und Schüler

Liebe Schülerin, lieber Schüler!

Deine Meinung ist gefragt! Auf den nächsten Seiten wirst Du zuerst ein paar Fragen über Dich finden und danach interessiert uns Deine Meinung zu der letzten Unterrichtsstunde.

Die Befragung ist freiwillig und anonym. Das bedeutet, dass niemand – weder Dein/e Lehrer/in noch Deine Eltern oder eine andere Person – erfährt, welche Antworten Du gegeben hast.

#### Beim Ausfüllen des Fragebogens achte bitte auf Folgendes:

- Lies Dir jede Aussage aufmerksam durch und wähle die Antwort aus, die Deiner Meinung nach am besten passt. Da es um Deine eigene Meinung geht, gibt es keine richtigen oder falschen Antworten.
- Kreuze das Kästchen an, das zu Deiner Antwort gehört.
- Kreuze bitte immer nur **ein** Kästchen in jeder Reihe an!
- Verwende bitte einen schwarzen oder blauen Stift zum Ausfüllen des Fragebogens und **keinen** Bleistift!

Dieses **Beispiel** zeigt Dir, wie es geht:

*Gib an, wie sehr Du den folgenden Aussagen zustimmst:*

	stimme nicht zu	stimme eher nicht zu	stimme eher zu	stimme zu
Was wir im Unterricht durchführen, verstehe ich meistens .....	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Wenn Du Deine Antwort ändern möchtest, dann machst Du das angekreuzte Kästchen ungültig, indem Du es ganz ausfüllst und stattdessen das Kästchen mit Deiner neuen Antwort ankreuzt:

Was wir im Unterricht durchführen, verstehe ich meistens .....	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
---	--------------------------	-------------------------------------	--------------------------	-------------------------------------

### Zuerst ein paar Fragen zu Deiner Person

A1 Bist Du ein Mädchen oder ein Junge?

☐ Mädchen

☐ Junge

A2 In welcher Klassenstufe bist Du?

☐ 7. Klasse

☐ 8. Klasse

☐ 9. Klasse

☐ 10. Klasse

A3 Welche Schulart besuchst Du?

☐ Hauptschule

☐ Realschule

☐ Gymnasium

☐ andere, nämlich: \_\_\_\_\_

A4 Welche Sprache sprichst Du selbst zu Hause am häufigsten?

☐ deutsch

☐ türkisch

☐ italienisch

☐ russisch

☐ andere, nämlich: \_\_\_\_\_

A5 Wie oft sprichst Du zu Hause Deutsch?

☐ Ich spreche zu Hause immer Deutsch.

☐ Ich spreche zu Hause manchmal Deutsch und manchmal eine andere Sprache.

☐ Ich spreche zu Hause niemals Deutsch.

A6 Welche Noten hast Du im letzten Zeugnis bekommen?

	sehr gut	gut	befriedigend	ausreichend	mangelhaft	ungenügend
	1	2	3	4	5	6
in Deutsch .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
in Mathematik .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
in Englisch .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
in Französisch .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
in Biologie .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
in Chemie .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
in Physik .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

A7 War die letzte Unterrichtsstunde anders als sonst?

☐ ja, völlig anders

☐ ja, etwas anders

☐ nein, genau so wie sonst

A8 Wie gern magst Du das Fach der letzten Unterrichtsstunde?

☐ sehr gern

☐ ziemlich gern

☐ mittel

☐ eher ungern

☐ sehr ungern

**Bitte gib nun Deine Meinung zur letzten Unterrichtsstunde an:**

		stimme nicht zu	stimme eher nicht zu	stimme eher zu	stimme zu
	<b>Klassenführung</b>				
1	Die Stunde hat pünktlich begonnen .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2	Niemand kam zu spät.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3	Niemand hat den Unterricht gestört .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4	In der Stunde ging es nur um Dinge, die mit dem Lernstoff zu tun haben .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5	Wir waren die ganze Stunde über aufmerksam und konzentriert.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6	Material oder Geräte (z.B. Overhead-Projektor) waren gleich verfügbar und einsatzbereit .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7	In der Stunde war es so ruhig, dass wir gut arbeiten konnten.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8	Es wurde keine Zeit für Dinge verwendet, die mit dem Lernstoff nichts zu tun haben .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9	Der Geräuschpegel war so niedrig, dass wir konzentriert arbeiten konnten.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10	Der Lehrer / die Lehrerin hatte die gesamte Klasse ständig im Blick .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11	Der Lehrer / die Lehrerin hat immer alles mitbekommen, was in der Klasse passiert ist .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12	Der Lehrer / die Lehrerin erklärte Aufgaben und Arbeitsblätter so, dass wir alle gleich wussten, was zu tun ist .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13	Wenn die Arbeitsform geändert wurde, wusste jeder gleich, was er/sie zu tun hat .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14	Wenn wir uns umgesetzt haben, ging kaum Zeit verloren ..	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15	Bestimmte Aufgaben sind von dazu eingeteilten Schüler/innen (z.B. Klassendienste) erledigt worden.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16	Einzelne Schüler/innen haben dem Lehrer / der Lehrerin geholfen (z.B. beim Austeilen von Arbeitsblättern, dem Abwischen der Tafel oder anderen Dingen).....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
17	Das Klassenzimmer war aufgeräumt .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
18	Die Klassenregeln waren allen bekannt.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
19	Der Lehrer / die Lehrerin achtete darauf, dass es gar nicht erst zu Störungen kam .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
20	Allen Schülern/innen war klar, wie man sich im Unterricht verhalten soll .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

		stimme nicht zu	stimme eher nicht zu	stimme eher zu	stimme zu
<b>Lernförderliches Klima</b>					
21	Der Umgangston zwischen dem Lehrer/ der Lehrerin und den Schüler/innen war freundlich .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
22	Wir haben während des Unterrichts öfter mal gelacht .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
23	Wir konnten entspannt lernen .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
24	Der Lehrer / die Lehrerin machte während des Unterrichts auch ab und zu Scherze .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
25	Mit Fehlern ist der Lehrer / die Lehrerin so umgegangen, dass man daraus etwas lernen konnte .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
26	Der Lehrer / die Lehrerin hat uns zu Beginn der Stunde freundlich begrüßt .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
27	Der Lehrer / die Lehrerin hat die Schüler/innen immer ausreden lassen, wenn sie dran waren .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
28	Der Lehrer / die Lehrerin hat Vorschläge und Ideen der Schüler/innen ernst genommen .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
29	Wenn der Lehrer / die Lehrerin eine Frage gestellt hat, hatten wir immer ausreichend Zeit zum Nachdenken .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
30	Der Lehrer / die Lehrerin hat uns unsere Fehler selbst korrigieren lassen .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
31	Nach Fehlern haben wir nicht nur erfahren, dass etwas falsch war, sondern auch warum .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
32	Wenn eine Antwort unvollständig war, hat uns der Lehrer / die Lehrerin Gelegenheit gegeben, die Antwort zu verbessern oder zu ergänzen .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
33	Der Lehrer / die Lehrerin hat freundlich reagiert, auch wenn jemand etwas Falsches gesagt hat .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
34	Der Lehrer / die Lehrerin hat die Atmosphäre durch humorvolle Bemerkungen aufgelockert .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Motivierung</b>					
35	Der Lehrer / die Lehrerin hat Verbindungen zu anderen Fächern hergestellt .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
36	Der Lehrer / die Lehrerin hat Beispiele aus dem Alltagsleben gegeben .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
37	Die Aufgaben waren abwechslungsreich .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
38	Die Arbeitsblätter waren interessant und haben uns Spaß gemacht .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
39	Der Lehrer / die Lehrerin hat uns erklärt, warum der in der Stunde behandelte Unterrichtsstoff für das spätere Lernen wichtig ist .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

5

		stimme nicht zu	stimme eher nicht zu	stimme eher zu	stimme zu
40	Der Lehrer / die Lehrerin hat uns gezeigt, dass wir den Unterrichtsstoff auch im Alltag gebrauchen können.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
41	Der Lehrer / die Lehrerin hat uns erklärt, dass der Unterrichtsstoff für das Lernen in anderen Fächern wichtig ist....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
42	Der Lehrer / die Lehrerin hat uns mit seinen / ihren Beispielen neugierig gemacht, mehr vom Stoff zu erfahren	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
43	Der Lehrer / die Lehrerin hat uns klar gemacht, dass es sich lohnt, sich für dieses Fach anzustrengen .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
44	Der Lehrer / die Lehrerin hat Aufgaben gewählt, die etwas mit meinen persönlichen Interessen zu tun haben.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
45	Bei guten Antworten sind die Schüler/innen von dem Lehrer / der Lehrerin gelobt worden .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
46	Der Wechsel zwischen verschiedenen Medien (z.B. Tafel, Poster, CD) hat die Stunde interessant gemacht .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
47	Die Schüler/innen waren während des Unterrichts konzentriert bei der Sache .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Klarheit und Verständlichkeit</b>					
48	Zu Unterrichtsbeginn hat der Lehrer / die Lehrerin erklärt, was wir in dieser Stunde lernen sollen .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
49	Mir war klar, was ich in dieser Stunde lernen sollte .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
50	Zu Unterrichtsbeginn hat der Lehrer / die Lehrerin darauf hingewiesen, was das Ziel dieser Stunde ist .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
51	Wenn andere Schüler gesprochen haben, konnte ich sie sehr gut verstehen.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
52	Der Lehrer / die Lehrerin hat so gesprochen, dass ich sie sehr gut verstehen konnte .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
53	Der Lehrer / die Lehrerin hat uns an Unterrichtsstoff erinnert, den wir früher einmal gelernt haben .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
54	Der Lehrer / die Lehrerin hat das, was besonders wichtig ist, deutlich hervorgehoben .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
55	Der Lehrer / die Lehrerin hat sich klar und deutlich ausgedrückt.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
56	Zu Beginn der Stunde hat der Lehrer / die Lehrerin Stoff aus vergangenen Stunden kurz wiederholt.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
57	Der Lehrer / die Lehrerin hat im Unterricht Lerninhalte aufgegriffen, die in den Hausaufgaben vorkamen .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
58	Gelegentliche Zusammenfassungen des Lehrers / der Lehrerin haben mir geholfen, den Lernstoff besser zu verstehen.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

6

		stimme nicht zu	stimme eher nicht zu	stimme eher zu	stimme zu
59	Der Lehrer / die Lehrerin hat am Ende der Stunde die wichtigsten Ergebnisse zusammengefasst .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
60	Es gab anschauliche Beispiele, die mir das Verstehen des Stoffes erleichtert haben.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
61	Der Lehrer / die Lehrerin hat darauf geachtet, dass wir die deutsche Sprache richtig verwenden .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
62	Der Lehrer / die Lehrerin hat darauf geachtet, dass wir uns klar und deutlich ausdrücken.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
63	Der Lehrer / die Lehrerin hat nachgeprüft, ob wir seine / ihre Erklärungen wirklich verstanden haben .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Aktivierung</b>					
64	Der Lehrer / die Lehrerin hat uns Hinweise gegeben, wie man besser lernen kann (z.B. zur Arbeitsplatzgestaltung) ..	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
65	Der Lehrer / die Lehrerin hat uns Hinweise zur Zeitplanung und Zeiteinteilung beim Lernen gegeben .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
66	Der Lehrer / die Lehrerin hat uns aufgefordert, Fragen zu stellen, wenn uns etwas nicht klar ist .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
67	Der Lehrer / die Lehrerin hat uns Tipps gegeben, wie wir Aufgaben selbstständig lösen können.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
68	Der Lehrer / die Lehrerin hat uns Aufgaben gegeben, die wir selbstständig lösen mussten.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
69	Es gab Aufgaben, die man auf unterschiedliche Weise lösen konnte .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
70	Wir haben unsere Arbeitsergebnisse selbst kontrolliert und korrigiert.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
71	Wir haben in der Stunde eigene Arbeitsergebnisse vorgestellt .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
72	Schüler/innen haben etwas an der Tafel oder am Overheadprojektor dargestellt.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
73	In der Unterrichtsstunde durften wir Schüler/innen uns gegenseitig aufrufen .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
74	In der Unterrichtsstunde stellten wir uns auch gegenseitig Fragen zur Unterrichtsthematik .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
75	In Kleingruppen haben Schüler/innen die Rolle des Lehrers (z.B. als Tutor) übernommen .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
76	Während der Stunde haben Schüler/innen von sich aus Vorschläge (z. B. zum Inhalt oder Ablauf der Stunde) gemacht.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
77	Die Schüler/innen haben den Lehrer / die Lehrerin durch Bedienung von Geräten unterstützt.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

7

		stimme nicht zu	stimme eher nicht zu	stimme eher zu	stimme zu
78	Wir haben verschiedene Lösungswege gesucht und sie vorgestellt .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
79	Der Lernstoff wurde durch Abbildungen oder Diagramme veranschaulicht (z.B. Mindmap) .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
80	Der Lehrer / die Lehrerin hat Fragen gestellt, die zum Nachdenken anregen .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
81	Der Lehrer / die Lehrerin hat uns selbst Aufgaben erfinden lassen .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
82	Ich habe mich während der Unterrichtsstunde mindestens einmal gemeldet .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
83	Ich bin während der Unterrichtsstunde mit mindestens einem Redebeitrag drangekommen .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## B.2 Standardisierter Interviewleitfaden zur Pilotierungsstudie

### Instruktion

- „Du bist ein Mithelfender, kein Befragter im ursprünglichen Sinne. Du leistest einen wichtigen Beitrag zur Verbesserung des Fragebogens. Dazu wollen wir nun zuerst die erste Seite bearbeiten. Lies dir die Seite bitte durch und gib mir bitte ein Zeichen, wenn du soweit bist.“
- SEITE 1
  - Zeit fürs Lesen : \_\_\_\_\_ min \_\_\_\_\_ sec
  - Als du eben die Seite gelesen hast, was ist dir dabei alles durch den Kopf gegangen?
    - im ersten allgemeinen Teil
    - beim zweiten Teil
    - beim letzten Teil
  - Was bedeutet für dich „letzte Unterrichtsstunde“?
  - Welche war das bei dir?
- SEITE 2
  - Zeit fürs Bearbeiten : \_\_\_\_\_ min \_\_\_\_\_ sec
  - Als du die Seite beantwortet hast, gab es Dinge, die dir unklar waren oder wo du Schwierigkeiten hattest bei der Beantwortung?
  - Als du eben die Frage „Welche Sprache sprichst du selbst zu Hause am häufigsten?“ beantwortet hast, was ist dir dabei alles durch den Kopf gegangen?
  - Als du eben die Frage „Wie oft sprichst du zu Hause Deutsch“ beantwortet hast, was ist dir dabei alles durch den Kopf gegangen?
  - Auf welches Zeugnis hast du deine Note bezogen?
  - Als du eben die Frage „War die letzte Unterrichtsstunde anders als sonst“ beantwortet hast, was ist dir dabei alles durch den Kopf gegangen?
  - Als du eben die Frage „Wie gern magst du das Fach der letzten Unterrichtsstunde“ beantwortet hast, was ist dir dabei alles durch den Kopf gegangen?
- SEITE 3 – KLASSENFÜHRUNG
  - Zeit fürs Bearbeiten : \_\_\_\_\_ min \_\_\_\_\_ sec
  - Als du die Seite beantwortet hast, gab es Dinge, die dir unklar waren oder wo du Schwierigkeiten hattest bei der Beantwortung?



- Stell dir vor, du müsstest jemandem erklären, was die vier Skalenpunkte bedeuten, was würdest du sagen?
  - Urteil
  - Sicherheit
- Was ist dir bei der Beantwortung der Frage alles durch den Kopf gegangen? – für Item 2, 3, 4, 5, 9, 10, 13, 16, 18, 19
- SEITE 4 – LERNFÖRDERLICHES KLIMA
  - Zeit fürs Bearbeiten : \_\_\_\_\_ min \_\_\_\_\_ sec
  - Als du die Seite beantwortet hast, gab es Dinge, die dir unklar waren oder wo du Schwierigkeiten hattest bei der Beantwortung?
  - Was ist dir bei der Beantwortung der Frage alles durch den Kopf gegangen? – für Item 22, 23, 24, 28, 30, 31, 32, 33, 34
- SEITE 4/5 – MOTIVIERUNG
  - Zeit fürs Bearbeiten : \_\_\_\_\_ min \_\_\_\_\_ sec
  - Als du die Seite beantwortet hast, gab es Dinge, die dir unklar waren oder wo du Schwierigkeiten hattest bei der Beantwortung?
  - Was ist dir bei der Beantwortung der Frage alles durch den Kopf gegangen? - für Item 35, 36, 41, 43, 44, 45, 46, 47
- SEITE 5 – KLARHEIT UND VERSTÄNDLICHKEIT
  - Zeit fürs Bearbeiten : \_\_\_\_\_ min \_\_\_\_\_ sec
  - Als du die Seite beantwortet hast, gab es Dinge, die dir unklar waren oder wo du Schwierigkeiten hattest bei der Beantwortung?
  - Was ist dir bei der Beantwortung der Frage alles durch den Kopf gegangen? – für Item 48, 49, 50, 52, 53, 54, 55, 57, 59, 62, 63
  - Präsentation von Item 48 und 49 in geänderter Reihenfolge → welche Konsequenzen im Frageverständnis ?
  - Item 51 – Stell dir vor, du müsstest jemandem erklären, was „sehr gut verstehen“ heißt, was würdest du sagen?
- SEITE 6 – AKTIVIERUNG
  - Zeit fürs Bearbeiten : \_\_\_\_\_ min \_\_\_\_\_ sec
  - Als du die Seite beantwortet hast, gab es Dinge, die dir unklar waren oder wo du Schwierigkeiten hattest bei der Beantwortung?

- Was ist dir bei der Beantwortung der Frage alles durch den Kopf gegangen? – für Item 64, 65, 66, 67, 68, 70, 71, 72, 73, 74, 75, 81, 82, 83,
  - Präsentation von Item 67 und 68 in geänderter Reihenfolge → welche Konsequenzen im Frageverständnis ?
  - Stell dir vor, du müsstest jemandem den Begriff **Arbeitsplatzgestaltung** erklären – was würdest du sagen?
  - Stell dir vor, du müsstest jemandem **Zeitplanung und -einteilung** erklären – was würdest du sagen?
  - Stell dir vor, du müsstest jemandem den Begriff **Tutor** erklären – was würdest du sagen?
  - Stell dir vor, du müsstest jemandem den Begriff **Redebeitrag** erklären – was würdest du sagen?
- **Noch ein paar allgemeine Fragen :**
- Hast du bereits Erfahrungen mit Interviews?
  - Wie fandest du das Interview?
  - Warum hast du mitgemacht?

## B.3 Itemvorlage zur Präferenzabfrage bezüglich der Lehrkraftbezeichnung

## Bezeichnung „der Lehrer“

	stimme nicht zu	stimme eher nicht zu	stimme eher zu	stimme zu
1 Der Lehrer hatte die gesamte Klasse ständig im Blick ...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2 Wenn der Lehrer eine Frage gestellt hat, hatten wir immer ausreichend Zeit zum Nachdenken.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3 Der Lehrer hat uns zu Beginn der Stunde freundlich begrüßt .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## Bezeichnung „die Lehrerin“

	stimme nicht zu	stimme eher nicht zu	stimme eher zu	stimme zu
1 Die Lehrerin hatte die gesamte Klasse ständig im Blick ..	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2 Wenn die Lehrerin eine Frage gestellt hat, hatten wir immer ausreichend Zeit zum Nachdenken.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3 Die Lehrerin hat uns zu Beginn der Stunde freundlich begrüßt .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## Bezeichnung „die Lehrperson“

	stimme nicht zu	stimme eher nicht zu	stimme eher zu	stimme zu
1 Die Lehrperson hatte die gesamte Klasse ständig im Blick .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2 Wenn die Lehrperson eine Frage gestellt hat, hatten wir immer ausreichend Zeit zum Nachdenken .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3 Die Lehrperson hat uns zu Beginn der Stunde freundlich begrüßt .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## Bezeichnung „die Lehrkraft“

	stimme nicht zu	stimme eher nicht zu	stimme eher zu	stimme zu
1 Die Lehrkraft hatte die gesamte Klasse ständig im Blick .	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2 Wenn die Lehrkraft eine Frage gestellt hat, hatten wir immer ausreichend Zeit zum Nachdenken.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3 Die Lehrkraft hat uns zu Beginn der Stunde freundlich begrüßt .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## Anhang C    Haupterhebung

### C.1.    Stichprobenrekrutierung

#### C.1.1    Plan zur Stichprobenrekrutierung

Abbildung C.1-1: *Rahmenplan zur Stichprobenakquise*

<b>Projekt "Unterrichtsdiagnostik"</b> <b>- Rahmenplan zur Stichprobenrekrutierung -</b> Stand: 15.10.2007			
<b>Zielgruppe*</b>	<b>Kontaktierungs- optionen</b>	<b>Wann?</b>	<b>Wer?</b>
<b>1. Moderatoren</b>			
Fachmoderatoren, Schul-Entwick- lungsmoderatoren, Trainer Führungs- kolleg, Gewaltmo- deratoren	1. Stufe 1) Veröffentlichung des Ausschreibungstextes auf der IfB-Homepage	voraus. möglich ab 17.10. (nach Fertig- stellung der End- version der Aus- schreibung)	veranlasst Frau Pikowsky
	2) Kontaktierung der Moderatorenverantwor- tlichen + Koreferenten externe Evaluation per eMail-Verteiler	ab 22.10. (nach Herbstferien)	Kontaktierung der Verantwortlichen der Moderatorengruppen über Frau Pikowsky, eMail-Versand über K. Piskol
	3) Anfrage Amtsblatt		
<b>2. Lehrkräfte allgemein</b>			
Fächer: Deutsch, Englisch, Franzö- sisch, Mathe- matik, Biologie, Chemie, Physik, Geographie	1. Stufe 1) Veröffentlichung des Ausschreibungstextes auf der IfB-Homepage	voraus. ab 17.10. (nach Fertigstellung der Endversion der Ausschreibung)	veranlasst Frau Pikowsky
	2. Stufe 2) Anschreiben der Schulen (über Schul- leitung) mit der Bitte die Ausschreibung ans Kollegium weiterzuleiten (Aushang an den für Lehrer zentralen Infor- mationsstellen)	in Abhängigkeit vom Rücklauf aus Stufe 1 - möglich ab 23.10.	Organisation + Versand übernimmt K. Piskol
	3. Stufe 3) optional: Vorstellung in Fortbildungsveranstal- tungen (sofern vorhanden)	in Abhängigkeit vom Rücklauf aus Stufe 2	

\* Klassenstufe 7-10 in Rheinland-Pfalz, keine berufsbildende Schulen

## C.1.2 Flyer zur Projektausschreibung



Institut für schulische Fortbildung  
und schulpyschologische Beratung  
des Landes Rheinland-Pfalz (IFB)



9. November 2007



Sehr geehrte Damen und Herren,

das IFB und die Universität Landau entwickeln derzeit einen **Fragebogen für Lehrkräfte zur Unterrichtsrückmeldung** aus der Schülerperspektive. Hintergrund ist, dass Lehrkräften ein praktikables Feedbackinstrument zur Einschätzung der Unterrichtsqualität zur Verfügung gestellt werden soll, welches als Grundlage zur Professionalisierung der eigenen Unterrichtstätigkeit herangezogen werden kann.

Um diesen Fragebogen zu überprüfen, soll eine konkrete Unterrichtsstunde aus Sicht der Lehrkraft und der Schüler/innen eingeschätzt werden – dazu möchten wir Sie um Ihre Unterstützung bitten!

Dieses Projekt wird vom Hochschulsonderprogramm "Wissen schafft Zukunft" der Landesregierung Rheinland-Pfalz finanziert und vom MBWJK nachhaltig unterstützt. Ziel dieses Gemeinschaftsprojektes "Unterrichtsdiagnostik" ist die Entwicklung wissenschaftlich fundierter und zugleich praktikabler Instrumente für die Unterrichtsrückmeldung und -entwicklung.

- Was?** Entwicklung eines Fragebogens zur Unterrichtsrückmeldung aus der Schülerperspektive als Feedbackinstrument für Lehrkräfte.
- Wer?** Mitmachen kann jede/r Lehrer/in, der/die in einem der Fächer Deutsch, Mathematik, Englisch, Französisch oder in einem naturwissenschaftlichen Fach in den Klassenstufen 7-10 unterrichtet (außer Berufsbildende Schulen und Förderschulen).
- Warum?** Sie haben die Möglichkeit, das neue Instrument kennen zu lernen und können einen Abgleich zwischen Ihrer Unterrichtseinschätzung und der Schülerwahrnehmung vornehmen. Sie erhalten eine ausführliche grafische Rückmeldung zum Profil Ihres Unterrichts.
- Wie?** Im Anschluss an eine Unterrichtsstunde soll diese sowohl von Ihnen als auch von Ihren Schülern/innen eingeschätzt werden (ca. 15 Minuten). Dafür erhalten Sie einen Klassensatz Fragebögen, den Sie entweder in einer Doppelstunde oder in Kooperation mit Ihrer Kollegin / Ihrem Kollegen der nachfolgenden Unterrichtsstunde von den Schülern/innen ausfüllen lassen können. Für den Rückversand der Fragebögen erhalten Sie einen frankierten Rückumschlag. Die Daten werden von uns zur Fragebogenentwicklung verwendet und nur Ihnen selbst rückgemeldet.
- Wann?** Das Projekt läuft bis Ende Januar 2008. Den genauen Erhebungstermin können Sie individuell festlegen.
- Kontakt?** Wenn Sie an dem Projekt teilnehmen möchten, senden Sie bitte das beiliegende Teilnahmeformular entweder per E-Mail, per Fax oder per Post an Kathleen Piskol. Sie können mich auch telefonisch unter ... erreichen.
- Zusatzinformation?** Wenn Sie darüber hinaus gehende Informationen zum Hintergrund wünschen, können Sie sich auch gern an Frau Dr. B. Pikowsky ([pikowsky@...](mailto:pikowsky@...)) oder an Herrn Prof. A. Helmke ([helmke@...](mailto:helmke@...)) wenden.

Wir freuen uns über Ihre Teilnahme und  
bedanken uns bereits im Voraus für Ihre Unterstützung!

Prof. Dr. Andreas Helmke  
Universität Koblenz-Landau

Kathleen Piskol  
Projektmitarbeiterin

Dr. Birgit Pikowsky  
IFB



Institut für schulische Fortbildung  
und schulpyschologische Beratung  
des Landes Rheinland-Pfalz (IFB)

Bitte senden Sie dieses **Teilnahmeformular**

entweder **per E-Mail** an:

kpiskol@...

oder **per Post** an:

Kathleen Piskol

oder **per Fax** an:

0621/...

### – Teilnahmeformular –

Wir freuen uns sehr über Ihre Bereitschaft zur Teilnahme an dem Projekt. Um mit Ihnen Kontakt aufzunehmen, möchten wir Sie bitten, uns folgende Angaben zukommen zu lassen:

	Name: <input type="text"/>	Vorname: <input type="text"/>
Schule	Name: <input type="text"/>	
	Adresse: <input type="text"/>	
Schulart	<input type="checkbox"/> HS <input type="checkbox"/> RS <input type="checkbox"/> Gym <input type="checkbox"/> IGS <input type="checkbox"/> RegS <input type="checkbox"/> DOS	
Fach	<input type="checkbox"/> Deutsch <input type="checkbox"/> Mathematik <input type="checkbox"/> Englisch <input type="checkbox"/> Französisch <input type="checkbox"/> Physik <input type="checkbox"/> Chemie <input type="checkbox"/> Biologie	
Klassenstufe	<input type="checkbox"/> 7. Klasse <input type="checkbox"/> 8. Klasse <input type="checkbox"/> 9. Klasse <input type="checkbox"/> 10. Klasse	
(voraus- sichtlicher) Erhebungs- termin	<input type="checkbox"/> Dezember 2007 <input type="checkbox"/> Januar 2008	
Kontaktaufnahme gewünscht:	<input type="checkbox"/> per E-Mail unter <input type="text"/>	<input type="checkbox"/> telefonisch unter <input type="text"/>

Vielen Dank!

Nach Erhalt Ihrer Rückmeldung melde ich mich unverzüglich bei Ihnen!

Mit freundlichen Grüßen

Kathleen Piskol  
Projektmitarbeiterin  
(Mobilnr: ...)

## C.1.3 Anschreiben für Lehrkräfte zur Befragungsdurchführung



Institut für schulische Fortbildung  
und schulpyschologische Beratung  
des Landes Rheinland-Pfalz (IFB)

Frau/Herrn ...		<b>Postadresse</b>
		... <b>Telefon:</b> ... <b>Telefax:</b> 0621/... kpiskol@...
		22. Januar 2008

**Projekt „Entwicklung eines Fragebogens zur Unterrichtsrückmeldung aus der Schülerperspektive als Feedbackinstrument zur Professionalisierung für Lehrkräfte“**

Liebe(r) Frau/Herr ... ,

noch einmal herzlichen Dank für Ihre schnelle und kompetente Unterstützung des Projekts „Unterrichtsrückmeldung aus der Schülerperspektive“. Anbei finden Sie Ihren Klassensatz Fragebögen für 27 Schüler/innen sowie einen Lehrerfragebogen für das Fach Physik/10. Klasse, wie telefonisch besprochen.

Der Fragebogen bezieht sich auf die letzte Unterrichtsstunde und sollte entweder in einer Doppelstunde (zu Beginn der zweiten Unterrichtsstunde) oder in Kooperation mit der Kollegin / dem Kollegen der (auf die einzuschätzende Unterrichtsstunde) folgenden Stunde erhoben werden. Die Erhebung dauert lediglich 15 Minuten.

Noch ein paar Anmerkungen zur Durchführung:

- Anbei finden Sie eine **Standardinstruktion** zur Durchführung der Befragung. Bitte lesen Sie diese den Schülern/innen vor oder teilen Sie ihnen die Informationen sinngemäß mit.
- Bitte teilen Sie die Fragebögen in der (bereits vorsortierten) Reihenfolge aus, da bei der Befragung **zwei Fragebogenversionen** eingesetzt werden, die sich hinsichtlich der Formulierung (jedoch nicht inhaltlich) unterscheiden. Bei der einen Version sollen die Schüler die Unterrichtsstunde aus der "Wir-Perspektive" für die gesamte Klasse einschätzen, bei der anderen erfolgt die Einschätzung der gleichen Aussagen aus der "Ich-Perspektive".
- Um die **Anonymität** zu gewährleisten, ist es wichtig, dass Sie vor Beginn der Befragung deutlich darauf hinweisen, dass Sie **nicht erfahren, was der/die einzelne Schüler/in geantwortet** hat, dass die Fragebögen nach dem Ausfüllen in einen Briefumschlag gesteckt werden und dieser verschlossen wird.

Bitte senden Sie alle Fragebogensätze nach der Erhebung zurück an: Kathleen Piskol, ... . Anbei finden Sie hierzu einen adressierten und frankierten Rückumschlag. Da ich die Erhebung Ende Januar abschließen möchte, wäre es schön, wenn Sie die Fragebögen zeitnah zurücksenden könnten.

Bei Rückfragen können Sie mich jederzeit gern per eMail unter [kpiskol@...](mailto:kpiskol@...) oder mobil unter ... erreichen.

Ich wünsche Ihnen eine gute Erhebung und  
sende freundliche Grüße aus Mannheim

Kathleen Piskol  
Projektmitarbeiterin

## C.1.4 Standardinstruktion für Lehrkräfte zur Befragungsdurchführung

- Ich möchte heute eine **Befragung** durchführen, bei der Eure **Meinung** zur **letzten Unterrichtsstunde** gefragt ist.
- Die Befragung ist **freiwillig** und **anonym**, d. h. niemand – weder ich *[wenn Kollegin / Kollege Befragung durchführt: Eure Lehrerin / Euer Lehrer]*, noch Eure Eltern oder eine andere Person erfährt, welche Antworten Ihr im Einzelnen gegeben habt. Die **Fragebögen** werden im Anschluss in einen **Briefumschlag gesteckt** und dieser wird **verschlossen**.
- In dem **Fragebogen** sollt Ihr die **letzte Unterrichtsstunde** *[bitte das entsprechende Unterrichtsfach nennen]* **anhand** einer **Reihe von Aussagen einschätzen** – dabei geht es um Eure persönliche Meinung, d. h. es gibt hier **keine richtigen oder falschen Antworten**.
- Ich **teile** Euch die **Fragebögen** nun **aus**, damit wir **anhand** eines **Beispiels** **durchgehen** können, **was** Ihr beim Ausfüllen **beachten** müsst *[Fragebögen in vorsortierter Reihenfolge austeilten]*.
- **Schaut** Euch bitte das **Beispiel an**:
  - **Lest** Euch **jede Aussage aufmerksam** durch (im Beispiel: Was wir im Unterricht durchführen, verstehe ich meistens) und **wählt** die **Antwort** aus, die Eurer Meinung nach **am besten passt**. Kreuzt dann bitte das Kästchen an, das zu der von Euch gewählten Antwort gehört. **ACHTUNG:** Ihr sollt **bei jeder Aussage** zu **zwei Fragen** Eure **Meinung** angeben:
    - 1) Wie schätzt Du die letzte Stunde ein?
    - 2) Wie sicher bist Du Dir bei Deiner Einschätzung?
- D. h. **in jeder Zeile** müssen **zwei Kästchen angekreuzt** sein.
- Wenn Ihr eine **Antwort ändern** möchtet, **füllt** das **falsch angekreuzte Kästchen** bitte ganz **aus** und **kreuzt** das **Kästchen an**, von dem Ihr glaubt, dass es **am besten passt**.
- Verwendet zum Ausfüllen bitte einen **schwarzen oder blauen Stift** und **keinen Bleistift**.
- Achtet bitte darauf, dass die Fragebogenseiten **auch** auf der **Rückseite bedruckt** sind und denkt daran, uns interessiert Eure **persönliche Meinung** – **arbeitet** also bitte **für Euch allein** *[wenn nötig Hinweis, dass der/die Nachbar/in jeweils eine andere Fragebogenversion hat]*.
- Gibt es noch **Fragen?** *[Fragen klären]*



## C.1.5 Informationsschreiben für die Schulleitung



Institut für schulische Fortbildung  
und schulpyschologische Beratung  
des Landes Rheinland-Pfalz (IFB)

9. November 2007

**Entwicklung eines Fragebogens zur Unterrichtsrückmeldung aus der Schülerperspektive als Feedbackinstrument zur Professionalisierung für Lehrkräfte**

Sehr geehrte Schulleiterin, sehr geehrter Schulleiter,

wir freuen uns, Ihnen mitteilen zu können, dass das Institut für schulische Fortbildung und schulpyschologische Beratung (IFB, Speyer) und die Universität Landau derzeit ein Projekt zur Entwicklung eines Fragebogens für Lehrkräfte zur Unterrichtsrückmeldung aus der Schülerperspektive realisieren, über das wir Sie mit diesem Schreiben informieren möchten.

Das innovative Vorhaben fokussiert die Idee, Schüler – aufgrund ihrer langjährigen Unterrichtserfahrungen – als Experten für Unterricht ernst zu nehmen und Lehrkräften diesbezüglich ein praktikables Feedbackinstrument zur Verfügung zu stellen, das sie zur Einschätzung ihres Unterrichts und als Hilfsmittel zur Professionalisierung ihrer Unterrichtstätigkeit heranziehen können.

Zur Überprüfung und Entwicklung dieses Fragebogens soll eine Befragung von Schülerinnen und Schülern der Klassenstufen 7-10 an allgemeinbildenden Schulen durchgeführt werden. Dabei wird eine konkrete Unterrichtsstunde von den Schülerinnen und Schülern sowie der jeweiligen Lehrkraft hinsichtlich verschiedener Unterrichtsmerkmale eingeschätzt. Die Befragung wird ca. 15 Minuten dauern und ist für den Zeitraum von Mitte November 2007 bis spätestens Januar 2008 angesetzt. Die erhobenen Daten werden zur Überprüfung des Fragebogens verwendet und nur der jeweiligen Lehrkraft grafisch auf aggregiertem Niveau rückgemeldet, d. h. auf Basis der rückgemeldeten Ergebnisse kann kein Rückschluss auf die befragten Schülerinnen und Schüler vorgenommen werden.

Dieses Projekt ist vom Ministerium für Bildung, Wissenschaft, Jugend und Kultur genehmigt und wird vom Hochschulsonderprogramm "Wissen schafft Zukunft" der Landesregierung Rheinland-Pfalz finanziert. Das Projekt ist Teil des Gemeinschaftsprojektes "Unterrichtsdiagnostik", welches die Entwicklung wissenschaftlich fundierter und zugleich praktikabler Instrumente zur Diagnostik, Rückmeldung und Entwicklung von Unterricht fokussiert.

Dieses Schreiben können Sie gern Ihrem Schulleiternbeirat zur Information weiterleiten.

Sofern Sie Rückfragen zur Durchführung der Befragung haben, können Sie sich jederzeit gern an die Projektmitarbeiterin Kathleen Piskol wenden – per eMail unter [kpiskol@...](mailto:kpiskol@...). Wenn Sie darüber hinausgehende Informationen zum Hintergrund wünschen, können Sie sich auch gern an Frau Dr. B. Pikowsky unter [pikowsky@...](mailto:pikowsky@...) oder an Herrn Prof. A. Helmke unter [helmke@...](mailto:helmke@...) wenden.



Mit freundlichen Grüßen


Dr. Birgit Pikowsky  
IFB

Kathleen Piskol  
Projektmitarbeiterin

Prof. Dr. Andreas Helmke  
Universität Koblenz-Landau

### C.1.6 Überprüfungsversion des Fragebogens (wir-Version mit männlicher Lehrkraftbezeichnung)



## Fragebogen für Schülerinnen und Schüler

---

Liebe Schülerin, lieber Schüler!

Deine Meinung zur letzten Unterrichtsstunde ist gefragt!

Die Befragung ist freiwillig und anonym. Das bedeutet, dass niemand – weder Dein Lehrer noch Deine Eltern oder eine andere Person – erfährt, welche Antworten Du gegeben hast.

---

**Beim Ausfüllen des Fragebogens achte bitte auf Folgendes:**

- Lies jede Aussage aufmerksam durch und wähle die Antwort aus, die Deiner Meinung nach am besten passt. Bei 1) sollst Du die Unterrichtsstunde anhand einer Reihe von Aussagen einschätzen. Bei 2) sollst Du angeben, wie sicher Du Dir bei Deiner Einschätzung bist. Es geht um Deine eigene Meinung, deshalb gibt es keine richtigen oder falschen Antworten.
- Kreuze das Kästchen an, das zu der gewählten Antwort gehört.
- Kreuze bitte immer nur ein Kästchen in jeder Spalte an!
- Verwende bitte einen schwarzen oder blauen Stift zum Ausfüllen des Fragebogens und keinen Bleistift!

Dieses **BEISPIEL** zeigt, wie es geht:



---

	1) Wie schätzt Du die letzte Stunde ein?				2) Wie sicher bist Du Dir bei Deiner Einschätzung?			
	stimme nicht zu	stimme eher nicht zu	stimme eher zu	stimme zu	sehr unsicher	eher unsicher	eher sicher	sehr sicher
Was wir im Unterricht durchführen, verstehe ich meistens.....	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Wenn Du Deine Antwort ändern möchtest, dann machst Du das angekreuzte Kästchen ungültig, indem Du es ganz ausfüllst und stattdessen das Kästchen mit Deiner neuen Antwort ankreuzt – also so:

---

Was wir im Unterricht durchführen, verstehe ich meistens.....	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
---	--------------------------	--------------------------	-------------------------------------	-------------------------------------	--------------------------	--------------------------	-------------------------------------	-------------------------------------



- 2 -

**Bitte gib nun Deine Meinung zur letzten Unterrichtsstunde an:**

[illegible]



	1) Wie schätzt Du die letzte Stunde ein?				2) Wie sicher bist Du Dir bei Deiner Einschätzung?			
	stimme nicht zu	stimme eher nicht zu	stimme eher zu	stimme zu	sehr unsicher	eher unsicher	eher sicher	sehr sicher
<b>Lernförderliches Klima</b>								
18	Der Umgangston zwischen dem Lehrer und den Schülern/innen war freundlich .....				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
19	Wir haben während des Unterrichts öfter mal gelacht .....				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
20	Wir konnten entspannt lernen.....				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
21	Mit Fehlern ist der Lehrer so umgegangen, dass wir daraus etwas lernen konnten .....				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
22	Der Lehrer hat die Klasse zu Beginn der Stunde freundlich begrüßt.....				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
23	Der Lehrer hat die Schüler/innen immer ausreden lassen, wenn sie dran waren .....				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
24	Der Lehrer hat Vorschläge und Ideen der Schüler/innen ernst genommen.....				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
25	Wenn der Lehrer eine Frage gestellt hat, hatten wir immer ausreichend Zeit zum Nachdenken .....				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
26	Der Lehrer machte während des Unterrichts auch ab und zu Scherze .....				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
27	Der Lehrer hat uns Fehler selbst korrigieren lassen .....				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
28	Nach Fehlern haben wir nicht nur erfahren, <u>dass</u> etwas falsch war, sondern auch <u>warum</u> .....				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
29	Bei unvollständigen Antworten hat uns der Lehrer Gelegenheit gegeben, die Antworten zu verbessern .....				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
30	Der Lehrer hat freundlich reagiert, auch wenn jemand etwas Falsches gesagt hat .....				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
31	Der Lehrer hat die Atmosphäre durch humorvolle Bemerkungen aufgelockert.....				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Motivierung</b>								
32	Der Lehrer hat Verbindungen zu anderen Fächern hergestellt.....				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
33	Der Lehrer hat Beispiele aus dem Alltagsleben gegeben.....				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
34	Die Aufgaben waren abwechslungsreich.....				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



- 4 -

[illegible]

- 5 -



		1) Wie schätzt Du die letzte Stunde ein?				2) Wie sicher bist Du Dir bei Deiner Einschätzung?			
		stimme nicht zu	stimme eher nicht zu	stimme eher zu	stimme zu	sehr unsicher	eher unsicher	eher sicher	sehr sicher
51	Der Lehrer hat im Unterricht Lerninhalte aufgegriffen, die in den Hausaufgaben vorkamen .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
52	Zu Unterrichtsbeginn hat der Lehrer darauf hingewiesen, was das Ziel dieser Stunde ist .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
53	Gelegentliche Zusammenfassungen des Lehrers haben uns geholfen, den Lernstoff besser zu verstehen .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
54	Die Mitschüler/innen haben so laut und deutlich gesprochen, dass wir sie gut verstehen konnten ...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
55	Es gab anschauliche Beispiele, die uns das Verstehen des Stoffes erleichtert haben .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
56	Der Lehrer hat darauf geachtet, dass wir die deutsche Sprache richtig verwenden .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
57	Der Lehrer hat am Ende der Stunde die wichtigsten Ergebnisse zusammengefasst .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
58	Der Lehrer hat darauf geachtet, dass wir uns klar und deutlich ausdrücken .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
59	Der Lehrer hat nachgeprüft, ob seine Erklärungen wirklich verstanden wurden .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
60	Uns war klar, was wir in dieser Stunde lernen sollten .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Aktivierung</b>									
61	Der Lehrer hat uns Hinweise gegeben, wie wir besser lernen können (z.B. zur Arbeitsplatzgestaltung) .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
62	Der Lehrer hat uns dazu aufgefordert, Fragen zu stellen, wenn uns etwas nicht klar ist .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
63	Der Lehrer hat uns Tipps gegeben, wie wir Aufgaben selbstständig lösen können .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
64	Wir haben unsere Arbeitsergebnisse selbst kontrolliert und korrigiert .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
65	Es gab Aufgaben, die wir auf unterschiedliche Weise lösen konnten .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
66	Der Lehrer hat uns Hinweise zur Zeitplanung und Zeiteinteilung beim Lernen gegeben .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>





- 6 -

[illegible]

## C.2 Stichprobenrelevante Angaben

## C.2.1 Stichprobenverteilung in Abhängigkeit von der Missinganzahl

Tabelle C.2-2: *Kreuztabellierte Häufigkeitsverteilung der Missings für die Items zur Unterrichtswahrnehmung (NMISS\_UNT) und Urteilssicherheit (NMISS\_US)*

NMISS_US							n
	0	1	2	3	4 bis 10	11 bis 18	
NMISS_UNT							
0	904	62	7	7	6	-	986
1	34	235	17	5	4	-	295
2	8	8	73	11	1	-	101
3	-	3	6	34	5	-	48
4 bis 10	5	5	3	3	59	2	77
11 bis 18	1	-	-	-	-	17	18
n	952	313	106	60	75	19	1525

## C.2.2 Häufigkeitsverteilung der Lehrkraftstichprobe hinsichtlich der Lehrberufsdauer

Tabelle C.2-3: *Häufigkeitsverteilung zur Lehrberufsdauer (in Jahren)*

Dauer Lehrtätigkeit	< 1 Jahr	1-4 Jahre	5-10 Jahre	11-20 Jahre	21-30 Jahre	> 30 Jahre	$\Sigma$
n	2	7	23	14	16	8	70

## C.2.3 Häufigkeitsverteilung der Lehrkräfte in Bezug auf die Dauer der Unterrichtstätigkeit in der befragten Klasse

Tabelle C.2-4: *Häufigkeitsverteilung zur Dauer der Unterrichtstätigkeit in der erhobenen Klasse*

Unterrichtsjahr	erstes Jahr	zweites Jahr	drittes Jahr	viertes Jahr	anderes Jahr	$\Sigma$
n	31	23	11	3	2 (3. und 5. Monat)	70



### C.3 Ergebnissrückmeldung

#### C.3.1 Anschreiben mit Informationen zum Umgang mit dem Ergebnisprofil



Institut für schulische Fortbildung  
und schulpyschologische Beratung  
des Landes Rheinland-Pfalz (IFB)

Universität Koblenz-Landau · Postfach 1864 · 55 008 Mainz

Frau/Herrn ...

7. Mai 2008

#### Ergebnissrückmeldung zum Projekt „Unterrichtsdiagnostik“

("Entwicklung eines Fragebogens zur Unterrichtsrückmeldung aus der Schülerperspektive als Feedbackinstrument zur Professionalisierung für Lehrkräfte")

Sehr geehrter Frau/Herr ... ,

Sie haben an dem Projekt „Unterrichtsrückmeldung aus der Schülerperspektive“ teilgenommen und eine konkrete Unterrichtsstunde anhand des Fragebogens selbst eingeschätzt und zugleich von Ihren Schülerinnen und Schülern einschätzen lassen. Im Folgenden finden Sie die Auswertungsergebnisse für Ihre Unterrichtsstunde (9. Klasse/Mathematik).

Vorab noch einige wichtige Hinweise zur Interpretation dieser Ergebnisse:

- Der Fragebogen erfasst die **fünf ausgewählten Unterrichtsmerkmale** Klassenführung, Lernförderliches Klima, Motivierung, Klarheit und Verständlichkeit sowie Aktivierung. Die Wirksamkeit dieser Merkmale für die Leistungssteigerung und die Entwicklung lern- sowie leistungsrelevanter Orientierungen ist in der Unterrichtsforschung nachgewiesen. Diese Merkmale sind als Eckpunkte für eine Orientierung zu verstehen – umso mehr Merkmale positiv ausgeprägt sind, umso höher ist die Lernerfolgswahrscheinlichkeit, aber es muss keineswegs jedes Merkmal einen Maximalwert aufweisen.
- Der Fragebogen leistet nur eine **Annäherung an die Komplexität des Unterrichtsgeschehens**. Er wurde mit dem Ziel der Einschätzung einer konkreten Unterrichtsstunde eingesetzt, so dass es durchaus vorkommen kann, dass bestimmte Fragen (z. B. Frage 36 „Die Arbeitsblätter haben den Schülern/innen Spaß gemacht“) für die Einschätzung Ihrer konkreten Unterrichtsstunde nicht relevant sind. Dies können nur Sie entscheiden und bei Ihrer Interpretation der Ergebnisse entsprechend berücksichtigen.
- Dieser Fragebogen dient der **Überprüfung der Fragen hinsichtlich Ihrer Eignung** zur Erfassung des jeweiligen Unterrichtsmerkmals. Deshalb ist er relativ umfangreich und enthält absichtlich einige inhaltlich sehr ähnliche Fragen (z. B. Frage 4 „In der Stunde ging es nur um Dinge, die mit dem Lernstoff zu tun haben“ und Frage 11 „Es wurde keine Zeit für Dinge verwendet, die mit dem Lernstoff nichts zu tun haben“) – mit dem Ziel, für die Erstellung der Endfassung die aussagekräftigsten Fragen auszuwählen.
- Der **Fragebogen ist fächerübergreifend** konzipiert, d. h., es kann vorkommen, dass bestimmte Fragen nicht exakt auf den jeweiligen Fachunterricht zutreffen (z. B. Frage 74 „Die Schüler/innen haben verschiedene Lösungswege gesucht und sie vorgestellt“).
- Ferner können bei einer Fragebogenerhebung **Messfehler** nie ausgeschlossen werden. Schwächen des Fragebogens aber auch äußere (z. B. Störgeräusche) und innere (z. B. Stimmungswechsel) situative Bedingungen können zu Über- oder Unterschätzungen führen.
- Aufgrund der aufgeführten Einschränkungen im Rahmen einer Fragebogenerhebung ist es wichtig, die Ergebnisse als eine **Anregung** zur Reflexion und Verbesserung des Unterrichts

zu sehen. Überprüfen Sie die Ergebnisse sorgfältig, bevor Sie Konsequenzen daraus ableiten. Hier kann ein Dialog mit Ihren Schülerinnen und Schülern sehr hilfreich sein.

- Insbesondere **Ergebnisse**, die Sie **nicht erwartet** haben und große **Diskrepanzen zwischen Ihrer Einschätzung und der Ihrer Schülerinnen und Schüler** bedürfen einer **inhaltlichen Interpretation**. Diese können nur Sie vornehmen, da Sie Experte für Ihre Klasse und die eingeschätzte Unterrichtsstunde sind. Lesen Sie die Frage, deren Ergebnis Sie erstaunt, nochmal genau durch und suchen Sie nach verschiedenen Begründungen für das Ergebnis bzw. die Diskrepanz – vielleicht helfen Ihnen dabei folgende Fragen:

- Gibt es in der Antwortverteilung einzelne Werte, die stark von den übrigen abweichen?
- Was haben Sie selbst angegeben und aus welchem Grund?
- Warum haben die Schülerinnen und Schüler diese Einschätzungen vorgenommen?
- Haben Sie und Ihre Schülerinnen und Schüler die Frage unterschiedlich verstanden?
- Haben die Schülerinnen und Schüler Ihre Handlungsintention erkannt?
- Haben Sie und Ihre Schülerinnen und Schüler unterschiedliche Vorstellungen und Erwartungen?

Überprüfen Sie Ihre Überlegungen, indem Sie entweder mit Ihren **Schülerinnen und Schülern darüber sprechen** oder aber bestimmte **Aspekte Ihres Unterrichts verändern** und die **Wirkung der Veränderung beobachten**.

Ab Seite 4 finden Sie pro Unterrichtsmerkmal eine Tabelle mit Ihren Auswertungsergebnissen mit den folgenden berechneten Werten, wobei die Antwortkategorien „stimme nicht zu“ mit 1, „stimme eher nicht zu“ mit 2, „stimme eher zu“ mit 3 und „stimme zu“ mit 4 kodiert sind:

- **Mittelwert der Schülereinschätzungen:** Summe aller Messwerte (Schülereinschätzungen) geteilt durch N (Anzahl der Schülereinschätzungen). Beantwortet die Frage: „Welches ist der durchschnittliche Messwert?“
- **Streuung der Schülereinschätzungen:** Maß für die Streuung einer Verteilung, das angibt, wie weit die Messwerte (Schülereinschätzungen) durchschnittlich vom Mittelwert entfernt sind. Beantwortet die Frage: „Wie typisch ist der durchschnittliche Messwert?“ Je mehr die einzelnen Messwerte von ihrem Mittelwert abweichen, desto größer ist die Streuung, d. h. desto unterschiedlicher wird Ihr Unterricht aus Schülersicht wahrgenommen.
- **Verteilung der Antworten:** Angabe der Anzahl der Schülereinschätzungen pro Antwortkategorie im Säulendiagramm.
- **% (3-4):** Prozentsatz der Schüler, die eine der beiden Antwortkategorien „stimme eher zu“ und „stimme zu“ angekreuzt haben. Beantwortet die Frage: „Wie groß ist der prozentuale Anteil an zustimmenden Schülereinschätzungen?“
- **N:** Anzahl der Schüler, die bei der jeweiligen Frage eine Einschätzung vorgenommen haben. Wichtig, um auf einen Blick zu sehen, ob es Fragen gibt, die von vielen ausgelassen wurden.

Das Verlaufsprofil (im rechten Teil der Abbildung) enthält die folgenden Werte:

- **Selbsteinschätzung (SE):** Ihre Einschätzung der Fragen in Gestalt eines Profils (sowie die dazugehörigen Schülereinschätzungen)

In der letzten Spalte finden Sie den Wert **„Differenz SE - M(S)“**. Dieser Wert ergibt sich aus der Differenz zwischen Ihrer Selbsteinschätzung und dem Mittelwert der Schülereinschätzungen und ist ein Maß dafür, wie stark Ihre Einschätzung mit der Ihrer Schülerinnen und Schüler übereinstimmt.

Zur Interpretation des Differenzwertes kann folgende **Faustregel** herangezogen werden: Unterschiede in der Größenordnung von etwa zwei Kategorien und mehr (Beispiel: Sie haben bei einer Frage den Wert 3 angegeben und der Mittelwert der Schülereinschätzungen beträgt 1) sind erheblich und sollten Anlass für Reflexion und gegebenenfalls eine Diskussion mit der Klasse sein.

Bei Rückfragen können Sie sich jederzeit gern an uns wenden. Bei Fragen zur Ergebnismeldung steht Ihnen die Projektmitarbeiterin, Frau Kathleen Piskol zur Verfügung – per E-Mail unter [kpiskol@...](mailto:kpiskol@...). Wenn Sie darüber hinausgehende Informationen wünschen, können Sie sich gern an Prof. Helmke wenden – per E-Mail unter [helmke@...](mailto:helmke@...).

In der Hoffnung, dass wir Ihnen mit dieser Rückmeldung hilfreiche Anregungen geben können, verbleiben wir

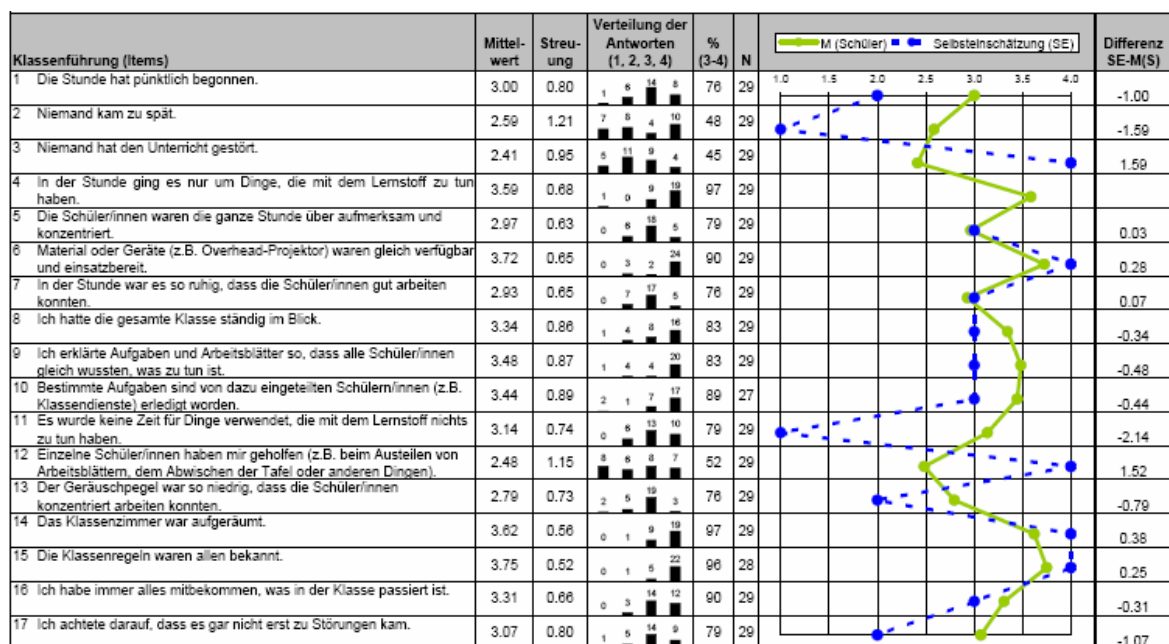
mit freundlichen Grüßen

Prof. Dr. Andreas Helmke  
Universität Koblenz-Landau

Kathleen Piskol  
Projektmitarbeiterin

Dr. Birgit Pikowsky  
IFB

### C.3.2 Beispiel eines Ergebnisprofils für das Unterrichtsmerkmal Klassenführung



## Anhang D Empirische Analysen zu den Unterrichtswahrnehmungsitems

### D.1. Deskriptivstatistische Befunde

#### D.1.1 Überblick zu den deskriptiven Itemkennwerten

Tabelle D.1-5: *Darstellung der deskriptiven Itemkennwerte zur Unterrichtswahrnehmung*

Items	Label	Item-form	Deskriptive Kennwerte					
			N	N <sub>miss</sub>	$\bar{x}_{insg.}$	$sd_{insg.}$	Schiefe	Exzess
SKF1	Stunde begann pünktlich		1517	8	3,27	0,91	-1,07	0,20
		ich-Form	768	5	3,28	0,89	-1,08	0,24
		wir-Form	749	3	3,25	0,92	-1,06	0,16
SKF2	niemand kam zu spät		1515	10	2,90	1,22	-0,53	-1,35
		ich-Form	769	4	2,90	1,22	-0,55	-1,32
		wir-Form	746	6	2,89	1,22	-0,52	-1,37
SKF3	niemand störte Unterricht		1507	18	2,47	1,03	-0,04	-1,15
		ich-Form	761	12	2,46	1,02	-0,03	-1,12
		wir-Form	746	6	2,49	1,04	-0,04	-1,18
SKF4	es ging nur um Dinge, die mit Lernstoff zu tun haben		1517	8	3,36	0,84	-1,24	0,81
		ich-Form	769	4	3,36	0,82	-1,13	0,51
		wir-Form	748	4	3,36	0,86	-1,34	1,06
SKF5	Schüler waren aufmerk-sam und konzentriert		1514	11	2,79	0,86	-0,41	-0,42
		ich-Form	767	6	2,95	0,84	-0,51	-0,29
		wir-Form	747	5	2,63	0,85	-0,33	-0,47
SKF6	Material/Geräte waren verfügbar/einsatzbereit		1466	59	3,21	1,05	-1,03	-0,31
		ich-Form	746	27	3,20	1,07	-1,00	-0,44
		wir-Form	720	32	3,22	1,03	-1,07	-0,16
SKF7	in Stunde war es so ruhig, dass Schüler gut arbeiten konnten		1518	7	2,82	0,90	-0,35	-0,66
		ich-Form	769	4	2,84	0,90	-0,39	-0,61
		wir-Form	749	3	2,80	0,90	-0,31	-0,70
SKF8	Lehrkraft hatte gesamte Klasse im Blick		1520	5	3,32	0,83	-1,08	0,43
		ich-Form	770	3	3,31	0,84	-1,04	0,31
		wir-Form	750	2	3,33	0,83	-1,13	0,58
SKF9	Lehrkraft erklärte Aufgaben so, dass alle wussten, was zu tun ist		1514	11	3,33	0,85	-1,14	0,54
		ich-Form	766	7	3,38	0,83	-1,23	0,76
		wir-Form	748	4	3,27	0,86	-1,05	0,37
SKF10	bestimmte Aufgaben wurden von eingeteilten Schülern erledigt		1434	91	2,81	1,10	-0,46	-1,12
		ich-Form	728	45	2,83	1,08	-0,49	-1,05
		wir-Form	706	46	2,79	1,12	-0,44	-1,19
SKF11	keine Zeit für Dinge verwendet, die nichts mit Lernstoff zu tun haben		1512	13	2,98	0,93	-0,56	-0,60
		ich-Form	766	7	2,99	0,93	-0,56	-0,61
		wir-Form	746	6	2,97	0,93	-0,56	-0,59
SKF12	einzelne Schüler helfen der Lehrkraft		1494	31	2,61	1,22	-0,17	-1,56
		ich-Form	761	12	2,63	1,21	-0,16	-1,53
		wir-Form	733	19	2,60	1,24	-0,17	-1,60

Items	Label	Item-form	Deskriptive Kennwerte					
			N	N <sub>miss</sub>	$\bar{x}_{insg.}$	$sd_{insg.}$	Schiefe	Exzess
SKF13	Geräuschpegel war so niedrig, dass Schüler konzentriert arbeiten konnten	ich-Form	1513	12	2,79	0,92	-0,36	-0,69
		wir-Form	764	9	2,82	0,93	-0,40	-0,69
		wir-Form	749	3	2,76	0,91	-0,33	-0,69
SKF14	Klassenzimmer war aufgeräumt	ich-Form	1515	10	3,23	0,89	-0,93	-0,04
		wir-Form	768	5	3,22	0,89	-0,91	-0,05
		wir-Form	747	5	3,24	0,90	-0,96	-0,03
SKF15	Klassenregeln waren Schülern bekannt	ich-Form	1495	30	3,45	0,88	-1,56	1,42
		wir-Form	760	13	3,51	0,88	-1,76	1,98
		wir-Form	735	17	3,38	0,88	-1,37	1,00
SKF16	Lehrkraft hat alles mitbekommen, was in Klasse passierte	ich-Form	1519	6	3,02	0,95	-0,61	-0,63
		wir-Form	771	2	3,02	0,92	-0,58	-0,61
		wir-Form	748	4	3,02	0,98	-0,64	-0,66
SKF17	Lehrkraft achtete darauf, dass es gar nicht erst zu Störungen kam	ich-Form	1521	4	2,94	0,90	-0,54	-0,45
		wir-Form	770	3	2,94	0,90	-0,56	-0,43
		wir-Form	751	1	2,94	0,89	-0,52	-0,47
SLK18	Lehrkraft verhielt sich freundlich zu Schülern	ich-Form	1521	4	3,60	0,71	-1,89	3,25
		wir-Form	770	3	3,62	0,69	-1,99	3,77
		wir-Form	751	1	3,57	0,73	-1,80	2,82
SLK19	Schüler haben öfter mal gelacht	ich-Form	1520	5	3,17	0,99	-0,93	-0,30
		wir-Form	771	2	3,00	1,07	-0,71	-0,82
		wir-Form	749	3	3,34	0,87	-1,14	0,38
SLK20	Schüler konnten entspannt lernen	ich-Form	1504	21	3,12	0,87	-0,76	-0,14
		wir-Form	758	15	3,07	0,86	-0,65	-0,27
		wir-Form	746	6	3,18	0,87	-0,88	0,04
SLK21	Lehrkraft ging mit Fehlern so um, dass Schüler daraus lernen konnten	ich-Form	1506	19	3,38	0,84	-1,31	1,06
		wir-Form	760	13	3,38	0,83	-1,31	1,06
		wir-Form	746	6	3,38	0,84	-1,33	1,06
SLK22	Lehrkraft begrüßte Schüler zu Stundenbeginn freundlich	ich-Form	1510	15	3,67	0,68	-2,34	5,33
		wir-Form	764	9	3,67	0,66	-2,27	5,16
		wir-Form	746	6	3,67	0,70	-2,40	5,45
SLK23	Lehrkraft hat Schüler immer ausreden lassen	ich-Form	1515	10	3,53	0,75	-1,70	2,51
		wir-Form	767	6	3,53	0,75	-1,66	2,37
		wir-Form	748	4	3,54	0,75	-1,74	2,68
SLK24	Lehrkraft hat Vorschläge der Schüler ernst genommen	ich-Form	1501	24	3,40	0,76	-1,21	1,07
		wir-Form	763	10	3,39	0,75	-1,12	0,84
		wir-Form	738	14	3,41	0,77	-1,30	1,30
SLK25	Schüler hatten bei Fragen ausreichend Zeit zum Nachdenken	ich-Form	1518	7	3,11	0,87	-0,74	-0,16
		wir-Form	770	3	3,10	0,85	-0,65	-0,32
		wir-Form	748	4	3,13	0,88	-0,83	0,01
SLK26	Lehrkraft machte ab und zu Scherze	ich-Form	1514	11	3,09	1,03	-0,82	-0,57
		wir-Form	766	7	3,15	1,02	-0,91	-0,43
		wir-Form	748	4	3,04	1,04	-0,75	-0,68
SLK27	Lehrkraft ließ Schüler Fehler selbst korrigieren	ich-Form	1502	23	3,05	0,90	-0,71	-0,27
		wir-Form	764	9	3,08	0,87	-0,70	-0,23
		wir-Form	738	14	3,01	0,93	-0,71	-0,33

Items	Label	Item-form	Deskriptive Kennwerte					
			N	N <sub>miss</sub>	$\bar{x}_{insg.}$	$sd_{insg.}$	Schiefe	Exzess
SLK28	Schüler haben nach Fehlern erfahren, dass und warum etwas falsch ist		1517	8	3,45	0,79	-1,46	1,61
		ich-Form	769	4	3,44	0,79	-1,42	1,46
		wir-Form	748	4	3,46	0,78	-1,50	1,79
SLK29	Lehrkraft gab Gelegenheit zur Verbesserung bei unvollständigen Antworten		1514	11	3,25	0,84	-0,98	0,33
		ich-Form	767	6	3,27	0,82	-0,96	0,29
		wir-Form	747	5	3,23	0,86	-1,00	0,34
SLK30	Lehrkraft freundlich, auch wenn Schüler etwas Falsches gesagt haben		1512	13	3,52	0,75	-1,61	2,19
		ich-Form	768	5	3,49	0,74	-1,46	1,84
		wir-Form	744	8	3,55	0,76	-1,76	2,62
SLK31	Lehrkraft lockerte Atmosphäre durch humorvolle Bemerkungen auf		1511	14	3,04	1,00	-0,72	-0,59
		ich-Form	765	8	3,04	1,00	-0,69	-0,65
		wir-Form	746	6	3,04	0,99	-0,75	-0,53
SM32	Lehrkraft stellte Verbindungen zu anderen Fächern her		1509	16	2,36	1,11	0,17	-1,32
		ich-Form	763	10	2,41	1,10	0,10	-1,32
		wir-Form	746	6	2,31	1,11	0,25	-1,29
SM33	Lehrkraft gab Beispiele aus Alltagsleben		1514	11	2,90	1,07	-0,54	-0,99
		ich-Form	765	8	2,91	1,07	-0,57	-0,97
		wir-Form	749	3	2,88	1,07	-0,52	-1,01
SM34	Schüler fanden Aufgaben abwechslungsreich		1517	8	2,87	1,01	-0,47	-0,89
		ich-Form	768	5	2,87	1,00	-0,47	-0,85
		wir-Form	749	3	2,87	1,02	-0,46	-0,93
SM35	Lehrkraft erklärte, warum Stoff fürs spätere Lernen wichtig ist		1515	10	2,44	1,08	0,08	-1,26
		ich-Form	769	4	2,45	1,08	0,07	-1,28
		wir-Form	746	6	2,43	1,07	0,09	-1,23
SM36	Arbeitsblätter machten Spaß		1450	75	2,54	1,07	-0,10	-1,24
		ich-Form	736	37	2,58	1,08	-0,14	-1,25
		wir-Form	714	38	2,51	1,06	-0,06	-1,21
SM37	Lehrkraft machte Schüler mit Beispielen neugierig, mehr zu erfahren		1508	17	2,57	0,98	-0,12	-0,98
		ich-Form	763	10	2,59	0,98	-0,12	-1,00
		wir-Form	745	7	2,55	0,97	-0,12	-0,96
SM38	Aufgaben hatten mit Interessen der Schüler zu tun		1516	9	2,48	1,04	0,03	-1,16
		ich-Form	767	6	2,49	1,04	0,02	-1,17
		wir-Form	749	3	2,46	1,04	0,04	-1,15
SM39	Lehrkraft erklärte, dass Stoff fürs Lernen in anderen Fächern wichtig ist		1508	17	2,32	1,06	0,24	-1,17
		ich-Form	763	10	2,34	1,07	0,22	-1,20
		wir-Form	745	7	2,31	1,06	0,27	-1,14
SM40	Lehrkraft zeigte, dass Schüler Stoff im Alltag gebrauchen können		1507	18	2,65	1,09	-0,21	-1,26
		ich-Form	762	11	2,67	1,10	-0,22	-1,27
		wir-Form	745	7	2,64	1,08	-0,20	-1,24
SM41	Lehrkraft machte klar, dass es sich lohnt, sich für dieses Fach anzustrengen		1511	14	2,91	1,02	-0,52	-0,90
		ich-Form	766	7	2,95	1,02	-0,57	-0,83
		wir-Form	745	7	2,88	1,03	-0,47	-0,96
SM42	Schüler wurden bei guten Antworten von Lehrkraft gelobt		1513	12	3,24	0,89	-1,00	0,13
		ich-Form	765	8	3,14	0,93	-0,81	-0,32
		wir-Form	748	4	3,34	0,85	-1,24	0,84

Items	Label	Item- form	Deskriptive Kennwerte					
			N	N <sub>miss</sub>	$\bar{x}_{insg.}$	$sd_{insg.}$	Schiefe	Exzess
SM43	Wechsel zwischen Medien hat Stunde interessant gemacht		1469	56	2,45	1,15	0,05	-1,43
		ich-Form	742	31	2,47	1,15	0,01	-1,43
		wir-Form	727	25	2,42	1,15	0,10	-1,42
SM44	Schüler waren während Unterricht aktiv bei der Sache		1513	12	2,97	0,87	-0,54	-0,38
		ich-Form	767	6	3,04	0,88	-0,60	-0,42
		wir-Form	746	6	2,91	0,85	-0,50	-0,31
SKV45	zu Unterrichtsbeginn erklärte Lehrkraft, was in Stunde gelernt werden soll		1520	5	2,76	1,11	-0,33	-1,23
		ich-Form	770	3	2,78	1,09	-0,36	-1,18
		wir-Form	750	2	2,73	1,12	-0,31	-1,29
SKV46	Lehrkraft so laut gesprochen, dass Schüler alles verstehen konnten		1520	5	3,65	0,66	-2,03	3,77
		ich-Form	770	3	3,63	0,68	-1,90	3,21
		wir-Form	750	2	3,68	0,65	-2,17	4,48
SKV47	Lehrkraft erinnerte an Stoff, den Schüler früher gelernt haben		1516	9	3,01	1,04	-0,68	-0,77
		ich-Form	769	4	2,93	1,05	-0,55	-0,94
		wir-Form	747	5	3,09	1,02	-0,82	-0,53
SKV48	Lehrkraft gab Hinweise zu besonders wichtigen Punkten		1518	7	3,15	0,89	-0,80	-0,18
		ich-Form	769	4	2,96	0,92	-0,54	-0,57
		wir-Form	749	3	3,33	0,82	-1,14	0,70
SKV49	Lehrkraft so klar ausgedrückt, dass Schüler Stoff verstanden haben		1517	8	3,33	0,83	-1,14	0,69
		ich-Form	768	5	3,33	0,84	-1,14	0,59
		wir-Form	749	3	3,32	0,81	-1,15	0,82
SKV50	zu Stundenbeginn wiederholte Lehrkraft Stoff aus vergangenen Stunden		1511	14	2,86	1,10	-0,48	-1,12
		ich-Form	764	9	2,87	1,10	-0,51	-1,10
		wir-Form	747	5	2,84	1,10	-0,45	-1,14
SKV51	Lehrkraft hat Lerninhalte aus Hausaufgaben aufgegriffen		1502	23	3,14	1,03	-0,94	-0,36
		ich-Form	756	17	3,17	1,02	-1,01	-0,19
		wir-Form	746	6	3,10	1,03	-0,87	-0,50
SKV52	Lehrkraft hat zu Unterrichtsbeginn darauf hingewiesen, was Ziel ist		1518	7	2,67	1,06	-0,20	-1,19
		ich-Form	770	3	2,66	1,04	-0,19	-1,14
		wir-Form	748	4	2,67	1,08	-0,21	-1,23
SKV53	Zusammenfassungen der Lehrkraft halfen, Lernstoff besser zu verstehen		1516	9	3,05	0,90	-0,68	-0,34
		ich-Form	769	4	3,05	0,90	-0,70	-0,27
		wir-Form	747	5	3,04	0,90	-0,66	-0,39
SKV54	Schüler so laut gesprochen, dass Mitschüler sie gut verstehen konnten		1517	8	3,01	0,93	-0,57	-0,64
		ich-Form	767	6	3,18	0,88	-0,81	-0,19
		wir-Form	750	2	2,84	0,95	-0,34	-0,87
SKV55	es gab Beispiele, die Verstehen erleichtert haben		1508	17	3,03	0,98	-0,64	-0,67
		ich-Form	764	9	3,02	0,98	-0,62	-0,75
		wir-Form	744	8	3,03	0,97	-0,67	-0,58
SKV56	Lehrkraft achtete darauf, dass Schüler deutsche Sprache richtig verwenden		1500	25	2,93	1,06	-0,55	-0,98
		ich-Form	762	11	2,90	1,08	-0,51	-1,05
		wir-Form	738	14	2,96	1,04	-0,58	-0,90
SKV57	Lehrkraft fasste am Stundenende Ergebnisse zusammen		1512	13	2,70	1,06	-0,25	-1,17
		ich-Form	766	7	2,73	1,05	-0,29	-1,12
		wir-Form	746	6	2,68	1,08	-0,21	-1,22

Items	Label	Item- form	Deskriptive Kennwerte					
			N	N <sub>miss</sub>	$\bar{x}_{insg.}$	$sd_{insg.}$	Schiefe	Exzess
SKV58	Lehrkraft achtete darauf, dass Schüler sich klar und deutlich ausdrücken		1511	14	3,09	0,90	-0,69	-0,42
		ich-Form	766	7	3,06	0,91	-0,67	-0,44
		wir-Form	745	7	3,11	0,89	-0,70	-0,38
SKV59	Lehrkraft prüfte, ob Schüler Erklärungen verstanden haben		1513	12	3,04	0,97	-0,70	-0,56
		ich-Form	767	6	2,91	0,99	-0,51	-0,81
		wir-Form	746	6	3,18	0,93	-0,93	-0,09
SKV60	Schülern war klar, was in dieser Stunde gelernt werden soll		1516	9	3,19	0,97	-0,96	-0,22
		ich-Form	769	4	3,22	0,98	-1,02	-0,13
		wir-Form	747	5	3,16	0,97	-0,90	-0,30
SA61	Lehrkraft gab Hinweise, wie Schüler besser lernen können		1498	27	2,30	1,10	0,24	-1,29
		ich-Form	758	15	2,28	1,09	0,27	-1,24
		wir-Form	740	12	2,32	1,11	0,21	-1,32
SA62	Lehrkraft forderte Schüler auf, Fragen zu stellen, wenn etwas unklar ist		1515	10	3,27	0,93	-1,06	0,05
		ich-Form	767	6	3,19	0,98	-0,93	-0,29
		wir-Form	748	4	3,35	0,88	-1,20	0,47
SA63	Lehrkraft gab Tipps, wie Schüler Aufgaben selbstständig lösen können		1504	21	2,81	1,00	-0,36	-0,96
		ich-Form	758	15	2,74	1,00	-0,25	-1,03
		wir-Form	746	6	2,88	1,00	-0,48	-0,85
SA64	Schüler haben Arbeitsergebnisse selbst kontrolliert und korrigiert		1511	14	2,74	1,03	-0,32	-1,06
		ich-Form	765	8	2,83	1,03	-0,42	-0,98
		wir-Form	746	6	2,66	1,03	-0,22	-1,10
SA65	es gab Aufgaben, die die Schüler auf unterschiedliche Weise lösen konnten		1513	12	2,94	1,03	-0,58	-0,85
		ich-Form	767	6	2,98	1,01	-0,65	-0,69
		wir-Form	746	6	2,89	1,06	-0,51	-0,99
SA66	Lehrkraft gab Hinweise zur Zeitplanung und -einteilung		1503	22	2,41	1,12	0,11	-1,36
		ich-Form	762	11	2,37	1,12	0,17	-1,34
		wir-Form	741	11	2,44	1,13	0,05	-1,38
SA67	Schüler haben eigene Arbeitsergebnisse vorgestellt		1512	13	2,60	1,15	-0,13	-1,41
		ich-Form	766	7	2,43	1,14	0,09	-1,40
		wir-Form	746	6	2,77	1,12	-0,36	-1,26
SA68	Lehrkraft gab Aufgaben, die die Schüler selbstständig lösen sollten		1521	4	3,35	0,92	-1,29	0,62
		ich-Form	772	1	3,31	0,94	-1,22	0,41
		wir-Form	749	3	3,39	0,90	-1,38	0,87
SA69	Schüler haben etwas an Tafel/am Overhead-Projektor dargestellt		1488	37	1,88	1,18	0,89	-0,86
		ich-Form	758	15	1,72	1,10	1,19	-0,15
		wir-Form	730	22	2,04	1,23	0,62	-1,30
SA70	Schüler durften andere Schüler aufrufen		1485	40	2,01	1,22	0,68	-1,21
		ich-Form	756	17	1,94	1,20	0,81	-1,01
		wir-Form	729	23	2,08	1,24	0,56	-1,37
SA71	Schüler haben von sich aus Vorschläge gemacht		1501	24	2,12	1,13	0,46	-1,23
		ich-Form	763	10	2,07	1,11	0,52	-1,14
		wir-Form	738	14	2,17	1,15	0,40	-1,32
SA72	Schüler haben Lehrkraft durch Bedienung technischer Geräte unterstützt		1480	45	1,81	1,06	0,98	-0,47
		ich-Form	753	20	1,65	1,00	1,29	0,31
		wir-Form	727	25	1,97	1,11	0,71	-0,94



Items	Label	Item- form	Deskriptive Kennwerte					
			N	N <sub>miss</sub>	$\bar{x}_{insg.}$	$sd_{insg.}$	Schiefe	Exzess
SA73	Schüler haben anderen Schülern Fragen zum Stoff gestellt		1507	18	2,15	1,16	0,42	-1,33
		ich-Form	763	10	2,00	1,13	0,63	-1,09
		wir-Form	744	8	2,31	1,18	0,23	-1,46
SA74	Schüler haben Lösungswege gesucht und vorgestellt		1505	20	2,33	1,14	0,19	-1,38
		ich-Form	763	10	2,09	1,11	0,53	-1,10
		wir-Form	742	10	2,58	1,12	-0,13	-1,34
SA75	Lernstoff durch Abbildungen veranschaulicht		1501	24	2,46	1,21	0,02	-1,56
		ich-Form	760	13	2,45	1,22	0,04	-1,58
		wir-Form	741	11	2,48	1,20	-0,01	-1,55
SA76	Lehrkraft stellte Fragen, die die Schüler zum Nachdenken anregen		1509	16	2,93	1,02	-0,57	-0,83
		ich-Form	768	5	2,89	1,03	-0,55	-0,87
		wir-Form	741	11	2,96	1,01	-0,60	-0,77
SA77	Lehrkraft hat Schüler Aufgaben erfinden lassen		1508	17	1,80	1,06	1,01	-0,38
		ich-Form	766	7	1,78	1,04	1,05	-0,26
		wir-Form	742	10	1,82	1,08	0,98	-0,50
SA78	Schüler sind während der Stunde mindestens einmal drangekommen		1519	6	3,02	1,13	-0,70	-0,99
		ich-Form	770	3	3,41	0,98	-1,51	0,94
		wir-Form	749	3	2,61	1,13	-0,14	-1,37

## D.1.2 Schiefe- und Kurtosiswerte ausgewählter ich- vs. wir-Items

Tabelle D.1-6: Angabe der Items und ihrer deskriptiven Kennwerte, die sich hinsichtlich ihrer Schiefe- und Exzesswerte im Vorzeichen unterscheiden.

Items	Label	Itemform	N	Schiefe	Exzess
SLK20	Schüler konnten entspannt lernen		1504	-0.76	-0.14
		ich-Version	758	-0.65	-0.27
		wir-Version	746	-0.88	0.04
SLK25	Schüler hatten bei Fragen ausreichend Zeit zum Nachdenken		1518	-0.74	-0.16
		ich-Version	770	-0.65	-0.32
		wir-Version	748	-0.83	0.01
SM42	Schüler wurden bei guten Antworten von Lehrkraft gelobt		1513	-1.00	0.13
		ich-Version	765	-0.81	-0.32
		wir-Version	748	-1.24	0.84
SKV48	Lehrkraft gab Hinweise zu besonders wichtigen Punkten		1518	-0.80	-0.18
		ich-Version	769	-0.54	-0.57
		wir-Version	749	-1.14	0.70
SA62	Lehrkraft forderte Schüler auf, Fragen zu stellen, wenn etwas unklar ist		1515	-1.06	0.05
		ich-Version	767	-0.93	-0.29
		wir-Version	748	-1.20	0.47
SA67	Schüler haben eigene Arbeitsergebnisse vorgestellt		1512	-0.13	-1.41
		ich-Version	766	0.09	-1.40
		wir-Version	746	-0.36	-1.26
SA72	Schüler haben Lehrkraft durch Bedienung technischer Geräte unterstützt		1480	0.98	-0.47
		ich-Version	753	1.29	0.31
		wir-Version	727	0.71	-0.94
SA74	Schüler haben Lösungswege gesucht und vorgestellt		1505	0.19	-1.38
		ich-Version	763	0.53	-1.10
		wir-Version	742	-0.13	-1.34
(SA75, neutrale Form)	Lernstoff durch Abbildungen veranschaulicht		1501	0,02	-1.56
		ich-Version	760	0.04	-1.58
		wir-Version	741	-0.01	-1.55
SA78	Schüler sind während der Stunde mindestens einmal drangekommen		1519	-0.70	-0.99
		ich-Version	770	-1.51	0.94
		wir-Version	749	-0.14	-1.37

## D.2 Item- und Reliabilitätsanalysen

## D.2.1 Überblick zu den Kennwerten der Item- und Reliabilitätsanalysen

Tabelle D.2-7: Darstellung der Item- und Reliabilitätswerte zur Unterrichtswahrnehmung

Items	Label	Item- form	Item- und Reliabilitätsanalyse-Kennwerte				
			N	P <sub>m</sub>	r <sub>ite</sub>	SK	Cronbachs α, wenn Item gelöscht
SKF1	Stunde begann pünktlich		1517	.76	.36	.20	.808
		ich-Form	768	.76	.37	.21	.811
		wir-Form	749	.75	.36	.19	.804
SKF2	niemand kam zu spät		1515	.63	.15	.06	.825
		ich-Form	769	.63	.16	.07	.829
		wir-Form	746	.63	.14	.06	.822
SKF3	niemand störte Unterricht		1507	.49	.43	.21	.803
		ich-Form	761	.49	.41	.20	.809
		wir-Form	746	.50	.46	.22	.798
SKF4	es ging nur um Dinge, die mit Lernstoff zu tun haben		1517	.79	.38	.23	.807
		ich-Form	769	.79	.37	.23	.812
		wir-Form	748	.79	.40	.23	.802
SKF5	Schüler waren aufmerk-sam und konzentriert		1514	.60	.49	.28	.801
		ich-Form	767	.65	.41	.24	.810
		wir-Form	747	.54	.58	.34	.792
SKF6	Material/Geräte waren verfügbar/einsatzbereit		1466	.74	.28	.13	.813
		ich-Form	746	.73	.29	.14	.817
		wir-Form	720	.74	.27	.13	.810
SKF7	in Stunde war es so ruhig, dass Schüler gut arbeiten konnten		1518	.61	.59	.33	.794
		ich-Form	769	.61	.60	.33	.798
		wir-Form	749	.60	.57	.32	.791
SKF8	Lehrkraft hatte gesamte Klasse im Blick		1520	.77	.54	.33	.798
		ich-Form	770	.77	.55	.33	.802
		wir-Form	750	.78	.54	.33	.794
SKF9	Lehrkraft erklärte Aufgaben so, dass alle wussten, was zu tun ist		1514	.78	.46	.27	.802
		ich-Form	766	.79	.46	.28	.807
		wir-Form	748	.76	.47	.27	.798
SKF10	bestimmte Aufgaben wurden von eingeteilten Schülern erledigt		1434	.60	.27	.12	.815
		ich-Form	728	.61	.28	.13	.818
		wir-Form	706	.60	.25	.11	.812
SKF11	keine Zeit für Dinge verwendet, die nichts mit Lernstoff zu tun haben		1512	.66	.43	.23	.804
		ich-Form	766	.66	.44	.24	.808
		wir-Form	746	.66	.41	.22	.801
SKF12	einzelne Schüler helfen der Lehrkraft		1494	.54	.26	.10	.818
		ich-Form	761	.54	.29	.12	.820
		wir-Form	733	.53	.22	.09	.817

Items	Label	Item- form	Item- und Reliabilitätsanalyse-Kennwerte				
			N	P <sub>m</sub>	r <sub>ite</sub>	SK	Cronbachs $\alpha$ , wenn Item gelöscht
SKF13	Geräuschpegel war so niedrig, dass Schüler konzentriert arbeiten konnten		1513	.60	.59	.32	.794
		ich-Form	764	.61	.60	.32	.798
		wir-Form	749	.59	.58	.31	.791
SKF14	Klassenzimmer war aufgeräumt		1515	.74	.39	.22	.806
		ich-Form	768	.74	.38	.22	.811
		wir-Form	747	.75	.39	.22	.802
SKF15	Klassenregeln waren Schülern bekannt		1495	.82	.37	.21	.807
		ich-Form	760	.84	.36	.20	.812
		wir-Form	735	.79	.38	.21	.803
SKF16	Lehrkraft hat alles mitbekommen, was in Klasse passierte		1519	.67	.62	.32	.792
		ich-Form	771	.67	.65	.35	.795
		wir-Form	748	.67	.59	.30	.789
SKF17	Lehrkraft achtete darauf, dass es gar nicht erst zu Störungen kam		1521	.65	.58	.33	.795
		ich-Form	770	.65	.63	.35	.796
		wir-Form	751	.65	.53	.30	.794
SLK18	Lehrkraft verhielt sich freundlich zu Schülern		1521	.87	.61	.43	.846
		ich-Form	770	.87	.60	.44	.837
		wir-Form	751	.86	.62	.42	.856
SLK19	Schüler haben öfter mal gelacht		1520	.72	.19	.09	.871
		ich-Form	771	.67	.21	.10	.864
		wir-Form	749	.78	.17	.10	.878
SLK20	Schüler konnten entspannt lernen		1504	.71	.58	.33	.847
		ich-Form	758	.69	.52	.30	.840
		wir-Form	746	.73	.64	.36	.854
SLK21	Lehrkraft ging mit Fehlern so um, dass Schüler daraus lernen konnten		1506	.79	.62	.37	.844
		ich-Form	760	.79	.59	.35	.836
		wir-Form	746	.79	.66	.39	.853
SLK22	Lehrkraft begrüßte Schüler zu Stundenbeginn freundlich		1510	.89	.47	.34	.853
		ich-Form	764	.89	.47	.36	.844
		wir-Form	746	.89	.47	.33	.863
SLK23	Lehrkraft hat Schüler immer ausreden lassen		1515	.84	.51	.34	.851
		ich-Form	767	.84	.49	.33	.842
		wir-Form	748	.84	.53	.35	.861
SLK24	Lehrkraft hat Vorschläge der Schüler ernst genommen		1501	.80	.56	.37	.848
		ich-Form	763	.80	.51	.35	.841
		wir-Form	738	.80	.59	.39	.857
SLK25	Schüler hatten bei Fragen ausreichend Zeit zum Nachdenken		1518	.70	.51	.30	.850
		ich-Form	770	.70	.51	.30	.841
		wir-Form	748	.71	.52	.30	.861
SLK26	Lehrkraft machte ab und zu Scherze		1514	.70	.51	.25	.852
		ich-Form	766	.72	.50	.24	.842
		wir-Form	748	.68	.52	.25	.862
SLK27	Lehrkraft ließ Schüler Fehler selbst korrigieren		1502	.68	.42	.23	.856
		ich-Form	764	.69	.44	.25	.845
		wir-Form	738	.67	.41	.22	.867

Items	Label	Item- form	Item- und Reliabilitätsanalyse-Kennwerte				
			N	P <sub>m</sub>	r <sub>ite</sub>	SK	Cronbachs α, wenn Item gelöscht
SLK28	Schüler haben nach Fehlern erfahren, dass und warum etwas falsch ist		1517	.82	.54	.34	.849
		ich-Form	769	.81	.51	.32	.841
		wir-Form	748	.82	.57	.36	.858
SLK29	Lehrkraft gab Gelegenheit zur Verbesserung bei unvollständigen Antworten		1514	.75	.55	.33	.849
		ich-Form	767	.76	.53	.33	.839
		wir-Form	747	.74	.56	.32	.859
SLK30	Lehrkraft freundlich, auch wenn Schüler etwas Falsches gesagt haben		1512	.84	.60	.40	.846
		ich-Form	768	.83	.61	.41	.836
		wir-Form	744	.85	.59	.39	.857
SLK31	Lehrkraft lockerte Atmosphäre durch humorvolle Bemerkungen auf		1511	.68	.61	.31	.844
		ich-Form	765	.68	.60	.30	.834
		wir-Form	746	.68	.62	.31	.855
SM32	Lehrkraft stellte Verbindungen zu anderen Fächern her		1509	.45	.43	.20	.857
		ich-Form	763	.47	.41	.19	.855
		wir-Form	746	.44	.46	.21	.861
SM33	Lehrkraft gab Beispiele aus Alltagsleben		1514	.63	.48	.23	.854
		ich-Form	765	.64	.46	.21	.852
		wir-Form	749	.63	.51	.24	.857
SM34	Schüler fanden Aufgaben abwechslungsreich		1517	.62	.59	.29	.848
		ich-Form	768	.62	.58	.29	.844
		wir-Form	749	.62	.61	.30	.851
SM35	Lehrkraft erklärte, warum Stoff fürs spätere Lernen wichtig ist		1515	.48	.53	.25	.851
		ich-Form	769	.48	.51	.24	.849
		wir-Form	746	.48	.55	.26	.855
SM36	Arbeitsblätter machten Spaß		1450	.51	.58	.27	.848
		ich-Form	736	.53	.53	.27	.844
		wir-Form	714	.50	.58	.28	.853
SM37	Lehrkraft machte Schüler mit Beispielen neugierig, mehr zu erfahren		1508	.52	.65	.33	.844
		ich-Form	763	.53	.65	.33	.840
		wir-Form	745	.52	.65	.33	.849
SM38	Aufgaben hatten mit Interessen der Schüler zu tun		1516	.49	.54	.26	.851
		ich-Form	767	.50	.52	.25	.848
		wir-Form	749	.49	.56	.27	.854
SM39	Lehrkraft erklärte, dass Stoff fürs Lernen in anderen Fächern wichtig ist		1508	.44	.56	.26	.849
		ich-Form	763	.45	.58	.27	.844
		wir-Form	745	.44	.55	.26	.855
SM40	Lehrkraft zeigte, dass Schüler Stoff im Alltag gebrauchen können		1507	.55	.57	.26	.849
		ich-Form	762	.56	.56	.25	.846
		wir-Form	745	.55	.58	.27	.853
SM41	Lehrkraft machte klar, dass es sich lohnt, sich für dieses Fach anzustrengen		1511	.64	.55	.27	.850
		ich-Form	766	.65	.53	.26	.848
		wir-Form	745	.63	.57	.28	.854
SM42	Schüler wurden bei guten Antworten von Lehrkraft gelobt		1513	.75	.41	.23	.858
		ich-Form	765	.71	.43	.23	.853
		wir-Form	748	.78	.41	.24	.862

Items	Label	Item- form	Item- und Reliabilitätsanalyse-Kennwerte				
			N	P <sub>m</sub>	r <sub>ite</sub>	SK	Cronbachs α, wenn Item gelöscht
SM43	Wechsel zwischen Medien hat Stunde interessant gemacht		1469	.48	.49	.21	.854
		ich-Form	742	.49	.51	.22	.849
		wir-Form	727	.47	.47	.20	.860
SM44	Schüler waren während Unterricht aktiv bei der Sache		1513	.66	.44	.25	.856
		ich-Form	767	.68	.43	.24	.853
		wir-Form	746	.64	.45	.26	.860
SKV45	zu Unterrichtsbeginn erklärte Lehrkraft, was in Stunde gelernt werden soll		1520	.59	.44	.20	.841
		ich-Form	770	.59	.46	.21	.831
		wir-Form	750	.58	.43	.19	.853
SKV46	Lehrkraft so laut gesprochen, dass Schüler alles verstehen konnten		1520	.88	.46	.34	.841
		ich-Form	770	.88	.44	.32	.832
		wir-Form	750	.89	.47	.37	.852
SKV47	Lehrkraft erinnerte an Stoff, den Schüler früher gelernt haben		1516	.67	.48	.23	.839
		ich-Form	769	.64	.46	.22	.830
		wir-Form	747	.70	.49	.24	.849
SKV48	Lehrkraft gab Hinweise zu besonders wichtigen Punkten		1518	.72	.53	.30	.836
		ich-Form	769	.65	.52	.28	.827
		wir-Form	749	.78	.57	.35	.847
SKV49	Lehrkraft so klar ausgedrückt, dass Schüler Stoff verstanden haben		1517	.78	.56	.34	.835
		ich-Form	768	.78	.55	.33	.826
		wir-Form	749	.77	.57	.35	.847
SKV50	zu Stundenbeginn wiederholte Lehrkraft Stoff aus vergangenen Stunden		1511	.62	.43	.20	.842
		ich-Form	764	.62	.41	.19	.834
		wir-Form	747	.61	.45	.20	.852
SKV51	Lehrkraft hat Lerninhalte aus Hausaufgaben aufgegriffen		1502	.71	.35	.17	.846
		ich-Form	756	.72	.36	.17	.836
		wir-Form	746	.70	.34	.16	.857
SKV52	Lehrkraft hat zu Unterrichtsbeginn darauf hingewiesen, was Ziel ist		1518	.56	.50	.24	.837
		ich-Form	770	.55	.347	.23	.829
		wir-Form	748	.56	.52	.24	.848
SKV53	Zusammenfassungen der Lehrkraft halfen, Lernstoff besser zu verstehen		1516	.68	.58	.32	.833
		ich-Form	769	.68	.56	.31	.825
		wir-Form	747	.68	.61	.34	.844
SKV54	Schüler so laut gesprochen, dass Mitschüler sie gut verstehen konnten		1517	.67	.25	.13	.850
		ich-Form	767	.73	.20	.11	.843
		wir-Form	750	.61	.32	.17	.858
SKV55	es gab Beispiele, die Verstehen erleichtert haben		1508	.68	.52	.27	.836
		ich-Form	764	.67	.50	.25	.828
		wir-Form	744	.68	.55	.28	.847
SKV56	Lehrkraft achtete darauf, dass Schüler deutsche Sprache richtig verwenden		1500	.64	.38	.18	.844
		ich-Form	762	.63	.38	.18	.835
		wir-Form	738	.65	.39	.19	.855
SKV57	Lehrkraft fasste am Stundenende Ergebnisse zusammen		1512	.57	.52	.25	.836
		ich-Form	766	.58	.51	.24	.827
		wir-Form	746	.56	.54	.25	.847

Items	Label	Item- form	Item- und Reliabilitätsanalyse-Kennwerte				
			N	P <sub>m</sub>	r <sub>ite</sub>	SK	Cronbachs $\alpha$ , wenn Item gelöscht
SKV58	Lehrkraft achtete darauf, dass Schüler sich klar und deutlich ausdrücken		1511	.70	.45	.25	.840
		ich-Form	766	.69	.44	.24	.831
		wir-Form	745	.70	.45	.25	.852
SKV59	Lehrkraft prüfte, ob Schüler Erklärungen verstanden haben		1513	.68	.55	.28	.835
		ich-Form	767	.64	.53	.27	.826
		wir-Form	746	.73	.58	.31	.846
SKV60	Schülern war klar, was in dieser Stunde gelernt werden soll		1516	.73	.54	.28	.835
		ich-Form	769	.74	.51	.26	.828
		wir-Form	747	.72	.57	.29	.846
SA61	Lehrkraft gab Hinweise, wie Schüler besser lernen können		1498	.43	.52	.23	.840
		ich-Form	758	.43	.53	.24	.842
		wir-Form	740	.44	.51	.23	.844
SA62	Lehrkraft forderte Schüler auf, Fragen zu stellen, wenn etwas unklar ist		1515	.76	.37	.20	.847
		ich-Form	767	.73	.36	.19	.842
		wir-Form	748	.78	.38	.22	.850
SA63	Lehrkraft gab Tipps, wie Schüler Aufgaben selbstständig lösen können		1504	.60	.52	.26	.841
		ich-Form	758	.58	.53	.26	.842
		wir-Form	746	.63	.51	.26	.845
SA64	Schüler haben Arbeitsergebnisse selbst kontrolliert und korrigiert		1511	.58	.38	.19	.846
		ich-Form	765	.61	.35	.17	.850
		wir-Form	746	.55	.44	.21	.847
SA65	es gab Aufgaben, die die Schüler auf unterschiedliche Weise lösen konnten		1513	.65	.44	.21	.844
		ich-Form	767	.66	.45	.22	.845
		wir-Form	746	.63	.44	.21	.848
SA66	Lehrkraft gab Hinweise zur Zeitplanung und -einteilung		1503	.47	.49	.22	.842
		ich-Form	762	.46	.48	.21	.844
		wir-Form	741	.48	.50	.22	.845
SA67	Schüler haben eigene Arbeitsergebnisse vorgestellt		1512	.53	.44	.19	.844
		ich-Form	766	.48	.49	.21	.843
		wir-Form	746	.59	.38	.17	.850
SA68	Lehrkraft gab Aufgaben, die die Schüler selbstständig lösen sollten		1521	.78	.31	.17	.849
		ich-Form	772	.77	.37	.20	.848
		wir-Form	749	.80	.24	.13	.855
SA69	Schüler haben etwas an Tafel/am Overhead-Projektor dargestellt		1488	.29	.48	.20	.842
		ich-Form	758	.24	.48	.22	.844
		wir-Form	730	.35	.47	.19	.846
SA70	Schüler durften andere Schüler aufrufen		1485	.34	.49	.20	.842
		ich-Form	756	.31	.49	.20	.844
		wir-Form	729	.36	.49	.20	.845
SA71	Schüler haben von sich aus Vorschläge gemacht		1501	.37	.53	.24	.840
		ich-Form	763	.36	.53	.24	.841
		wir-Form	738	.39	.53	.23	.843
SA72	Schüler haben Lehrkraft durch Bedienung technischer Geräte unterstützt		1480	.27	.53	.25	.840
		ich-Form	753	.22	.55	.28	.841
		wir-Form	727	.32	.51	.23	.844

Items	Label	Item- form	Item- und Reliabilitätsanalyse-Kennwerte				
			N	P <sub>m</sub>	r <sub>ite</sub>	SK	Cronbachs $\alpha$ , wenn Item gelöscht
SA73	Schüler haben anderen Schülern Fragen zum Stoff gestellt		1507	.38	.47	.20	.843
		ich-Form	763	.33	.43	.19	.846
		wir-Form	744	.44	.50	.21	.845
SA74	Schüler haben Lösungswege gesucht und vorgestellt		1505	.44	.53	.23	.840
		ich-Form	763	.36	.54	.24	.841
		wir-Form	742	.53	.53	.24	.843
SA75	Lernstoff durch Abbildungen veranschaulicht		1501	.49	.38	.16	.847
		ich-Form	760	.48	.35	.15	.850
		wir-Form	741	.49	.41	.17	.849
SA76	Lehrkraft stellte Fragen, die die Schüler zum Nachdenken anregen		1509	.64	.48	.23	.842
		ich-Form	768	.63	.50	.24	.843
		wir-Form	741	.65	.46	.23	.846
SA77	Lehrkraft hat Schüler Aufgaben erfinden lassen		1508	.27	.48	.23	.842
		ich-Form	766	.26	.48	.23	.844
		wir-Form	742	.27	.48	.22	.845
SA78	Schüler sind während der Stunde mindestens einmal drangekommen		1519	.67	.28	.12	.851
		ich-Form	770	.80	.28	.14	.852
		wir-Form	749	.54	.41	.18	.849



## D.3 Faktorenanalytische Auswertung

## D.3.1 Faktorenstruktur exploratorischer Faktorenlösungen

Tabelle D.3-1: Überblick zu exploratorischen Faktorenlösungen

	Interpretation der Faktorenlösungen (HKA, Oblimin)		
	4er Lösung	5er Lösung	6er Lösung
<b>Gesamt- stichprobe</b>	F1: LK	<b>F1: LK</b>	F1: LK
	F2: A	<b>F2: A</b>	<b>F2: A</b>
	<b>F3: KF</b>	<b>F3: KF</b>	<b>F3: KF</b>
	F4: M, KV	F4: M, KV F5: KV (wenige Items)	<b>F4: M</b> F5: KV F6: KF, M, KV
<b>ich-Version</b>	F1: LK	<b>F1: LK</b>	F1: LK
	F2: A, KV	<b>F2: A</b>	<b>F2: A</b>
	<b>F3: KF</b>	<b>F3: KF</b>	<b>F3: KF</b>
	F4: M, KV, A	F4: M, KV F5: M, KV, A	<b>F4: M</b> F5: KV, A F6: KF, M, KV, A
<b>wir-Version</b>	F1: LK, KV	<b>F1: LK</b>	F1: LK, KV
	F2: A	<b>F2: A</b>	<b>F2: A</b>
	<b>F3: KF</b>	<b>F3: KF</b>	<b>F3: KF</b>
	F4: M, KV	F4: M F5: KV	<b>F4: M</b> F5: KV, A F6: M, KV, A

*Anmerkung:* Die sich anhand von Markiervariablen eindeutig ergebenden Faktoren sind hervorgehoben, dabei ist KF = Klassenführung, A = Aktivierung, LK = Lernförderliches Klima, M = Motivierung und KV = Klarheit und Verständlichkeit.

## D.3.2 Faktorladungsmatrix der 5-Faktorenlösung

Tabelle D.3-8: Faktorladungsmuster für die 5-Faktorenlösung (HKA, Oblimin) für die ich-/wir-Version und die Gesamtstichprobe

Item	Version	Unrotierte Lösung (F1)	Interpretation der Faktorenlösung					Kommunalität
			Gesamtstichprobe					
			KV	A	KF	-M	LK	
			ich-Version					
			A, M, KV	A	KF	M, KV	-LK	
wir-Version								
KV	A	KF	-M	LK				
SKF1	insg.	.36			.39		.20	
	ich	.34			.44		.24	
	wir	.38			.34		.19	
SKF2	insg.						.03	
	ich						.05	
	wir						.02	
SKF3	insg.	.35			.63		.40	
	ich	.31			.76		.50	
	wir	.39			.63		.40	
SKF4	insg.	.35			.39		.22	
	ich	.33			.33		.19	
	wir	.38			.41		.27	
SKF5	insg.	.45			.70		.50	
	ich	.42	.40		.43		.45	
	wir	.51			.73		.56	
SKF6	insg.						.10	
	ich						.09	
	wir	.31					.12	
SKF7	insg.	.47			.74		.55	
	ich	.48			.79		.59	
	wir	.46			.79		.56	
SKF8	insg.	.54			.42		.39	
	ich	.53			.48		.43	
	wir	.55			.42		.39	
SKF9	insg.	.55			.34		.38	
	ich	.55	.34				.40	
	wir	.55			.36	.36	.38	
SKF10	insg.						.10	
	ich						.08	
	wir	.30					.12	
SKF11	insg.	.35			.52		.29	
	ich	.35			.49		.29	
	wir	.35			.53		.32	
SKF12	insg.	.35					.14	
	ich	.36					.15	
	wir	.34					.15	

Item	Version	Unrotierte Lösung (F1)	Interpretation der Faktorenlösung					Kommunalität
			Gesamtstichprobe					
			KV	A	KF	-M	LK	
			ich-Version					
			A, M, KV	A	KF	M, KV	-LK	
wir-Version								
KV	A	KF	-M	LK				
SKF13	insg.	.50			<b>.74</b>		.56	
	ich	.53			<b>.76</b>		.59	
	wir	.49			<b>.74</b>		.58	
SKF14	insg.	.38			<b>.33</b>		.21	
	ich	.37			<b>.33</b>		.22	
	wir	.39			<b>.34</b>		.21	
SKF15	insg.	.39			<b>.30</b>		.22	
	ich	.36					.20	
	wir	.42					.23	
SKF16	insg.	.59			<b>.56</b>		.48	
	ich	.61			<b>.55</b>		.53	
	wir	.58			<b>.57</b>		.47	
SKF17	insg.	.55			<b>.57</b>		.44	
	ich	.58			<b>.59</b>		.51	
	wir	.53			<b>.60</b>		.45	
SLK18	insg.	.54				<b>.60</b>	.50	
	ich	.54				<b>-.54</b>	.49	
	wir	.55				<b>.63</b>	.52	
SLK19	insg.	.62			-.34	<b>.44</b>	.24	
	ich	.60				<b>-.48</b>	.22	
	wir	.65			-.40	<b>.38</b>	.26	
SLK20	insg.	.63			.45	.38	.50	
	ich	.62			.47		.47	
	wir	.64			.41	.47	.55	
SLK21	insg.	.43				<b>.49</b>	.50	
	ich	.39				<b>-.36</b>	.47	
	wir	.48				<b>.58</b>	.53	
SLK22	insg.	.51				<b>.50</b>	.29	
	ich	.49				<b>-.49</b>	.28	
	wir	.53				<b>.50</b>	.31	
SLK23	insg.	.55				<b>.44</b>	.36	
	ich	.49				<b>-.39</b>	.34	
	wir	.60				<b>.45</b>	.38	
SLK24	insg.	.55				<b>.54</b>	.42	
	ich	.53				<b>-.48</b>	.38	
	wir	.57				<b>.55</b>	.47	
SLK25	insg.	.46				<b>.34</b>	.34	
	ich	.44					.33	
	wir	.47				<b>.36</b>	.35	
SLK26	insg.	.40				<b>.62</b>	.50	
	ich	.42				<b>-.62</b>	.44	
	wir	.39			-.31	<b>.60</b>	.51	
SLK27	insg.	.53				<b>.42</b>	.23	
	ich	.50				<b>-.42</b>	.27	
	wir	.55				<b>.40</b>	.23	

Item	Version	Unrotierte Lösung (F1)	Interpretation der Faktorenlösung					Kommunalität	
			Gesamtstichprobe						
			KV	A	KF	-M	LK		
			ich-Version						
			A, M, KV	A	KF	M, KV	-LK		
wir-Version									
KV	A	KF	-M	LK					
SLK28	insg.	.54					.51	.39	
	ich	.56					-.39	.36	
	wir	.53					.58	.44	
SLK29	insg.	.52					.45	.37	
	ich	.52					-.36	.37	
	wir	.53					.52	.38	
SLK30	insg.	.58					.66	.49	
	ich	.56					-.61	.50	
	wir	.61					.65	.49	
SLK31	insg.	.39					.64	.55	
	ich	.38					-.63	.52	
	wir	.40					.60	.54	
SM32	insg.	.45	-.36				-.47	.38	
	ich	.43						.28	
	wir	.46					-.52	.36	
SM33	insg.	.61					-.51	.40	
	ich	.61					.34	.29	
	wir	.60					-.55	.43	
SM34	insg.	.52					-.44	.41	
	ich	.52					.41	.41	
	wir	.53					-.46	.44	
SM35	insg.	.62					-.59	.40	
	ich	.60					.63	.44	
	wir	.63					-.52	.42	
SM36	insg.	.65					-.41	.40	
	ich	.65					.48	.41	
	wir	.64					-.34	.43	
SM37	insg.	.52					-.52	.49	
	ich	.51					.48	.46	
	wir	.52					-.48	.49	
SM38	insg.	.46					-.46	.35	
	ich	.48	.35				.43	.39	
	wir	.45					-.45	.37	
SM39	insg.	.48					-.71	.49	
	ich	.49					.67	.46	
	wir	.47					-.60	.46	
SM40	insg.	.55					-.71	.47	
	ich	.53					.70	.43	
	wir	.57					-.64	.48	
SM41	insg.	.50					-.57	.42	
	ich	.51					.55	.37	
	wir	.49					-.51	.47	
SM42	insg.	.49						.39	.30
	ich	.50						-.31	.32
	wir	.48						.44	.32

Item	Version	Unrotierte Lösung (F1)	Interpretation der Faktorenlösung					Kommunalität
			Gesamtstichprobe					
			KV	A	KF	-M	LK	
			ich-Version					
			A, M, KV	A	KF	M, KV	-LK	
wir-Version								
			KV	A	KF	-M	LK	
SM43	insg.	.55				<b>-.45</b>		.32
	ich	.51				<b>.46</b>		.33
	wir	.61				<b>-.40</b>		.31
SM44	insg.	.44			.48			.41
	ich	.44	.63					.57
	wir	.44			.42			.42
SKV45	insg.	.51						.24
	ich	.48				.39		.23
	wir	.55	<b>.37</b>					.26
SKV46	insg.	.49					.45	.43
	ich	.47					-.30	.35
	wir	.51					.53	.49
SKV47	insg.	.55						.27
	ich	.52				.36		.25
	wir	.60	.33				.33	.34
SKV48	insg.	.61					.31	.34
	ich	.61				.36		.28
	wir	.60					.38	.42
SKV49	insg.	.44	<b>.36</b>					.45
	ich	.42	.38			.32		.47
	wir	.45	.34				.33	.45
SKV50	insg.	.37				-.35		.24
	ich	.39				.45		.24
	wir	.35						.25
SKV51	insg.	.46						.16
	ich	.44						.17
	wir	.47	<b>.31</b>					.17
SKV52	insg.	.58	<b>.37</b>					.29
	ich	.58				.49		.26
	wir	.59	<b>.49</b>					.36
SKV53	insg.	.32	<b>.33</b>			-.33		.40
	ich					.50		.39
	wir	.39	<b>.38</b>					.42
SKV54	insg.	.58			.30			.16
	ich	.56	<b>.52</b>					.27
	wir	.61			.37			.22
SKV55	insg.	.42						.34
	ich	.41				.38		.34
	wir	.43						.37
SKV56	insg.	.53				-.32		.25
	ich	.53				.42		.23
	wir	.54				-.35		.29
SKV57	insg.	.46				-.31		.31
	ich	.47				.38		.29
	wir	.45	<b>.32</b>					.34

Item	Version	Unrotierte Lösung (F1)	Interpretation der Faktorenlösung					Kommunalität
			Gesamtstichprobe					
			KV	A	KF	-M	LK	
			ich-Version					
			A, M, KV	A	KF	M, KV	-LK	
wir-Version								
KV	A	KF	-M	LK				
SKV58	insg.	.60	.34					.29
	ich	.60				.32		.26
	wir	.61	.37					.31
SKV59	insg.	.52				.45		.39
	ich	.53	.37					.38
	wir	.52						.43
SKV60	insg.	.49	.46					.41
	ich	.50	.31			.40		.36
	wir	.49	.54					.43
SA61	insg.	.48				-.55		.45
	ich	.47				.59		.44
	wir	.50	.32			-.46		.47
SA62	insg.	.56					.33	.30
	ich	.58						.25
	wir	.55	.31				.34	.33
SA63	insg.	.37	.34			-.31		.42
	ich	.35				.55		.42
	wir	.40	.45					.43
SA64	insg.	.45		.32				.23
	ich	.47						.17
	wir	.45		.40				.28
SA65	insg.	.45		.31				.26
	ich	.44						.26
	wir	.47		.36				.26
SA66	insg.	.38				-.36		.34
	ich	.40				.44		.32
	wir	.37	.30			-.31		.37
SA67	insg.	.40		.53				.37
	ich	.46	.48	.33				.39
	wir	.33		.49				.33
SA68	insg.	.30						.23
	ich		.35					.27
	wir	.33						.16
SA69	insg.	.36		.71				.45
	ich	.35		.64				.43
	wir	.36		.70				.44
SA70	insg.	.36		.60				.39
	ich	.35		.55				.38
	wir	.38		.63				.40
SA71	insg.	.35		.63				.42
	ich	.32		.59				.42
	wir	.38		.64				.43
SA72	insg.	.35		.65				.46
	ich			.64				.50
	wir	.40		.62				.43

Item	Version	Unrotierte Lösung (F1)	Interpretation der Faktorenlösung					Kommunalität
			Gesamtstichprobe					
			KV	A	KF	-M	LK	
			ich-Version					
			A, M, KV	A	KF	M, KV	-LK	
wir-Version								
KV	A	KF	-M	LK				
SA73	insg.	.40	.63			.38		
	ich	.40	.56			.34		
	wir	.41	.62			.39		
SA74	insg.	.38	.73			.49		
	ich	.37	.61			.49		
	wir	.39	.69			.44		
SA75	insg.	.54	.35			.23		
	ich	.52	.31			.23		
	wir	.55	.40			.27		
SA76	insg.	.36				.32		
	ich	.30				.34		
	wir	.35				.33		
SA77	insg.		.60			.41		
	ich	.35	.64			.44		
	wir	.43	.61			.39		
SA78	insg.					.15		
	ich	.50				.31		
	wir		.30			.23		
Gesamtstichprobe								
Eigenwert		17,41	4,16	2,34	1,91	1,75		
% Varianz		22,32	5,33	3,00	2,44	2,24		
ich-Version								
Eigenwert		2,01	4,27	1,83	16,86	2,32		
% Varianz		2,58	5,48	2,35	21,61	2,98		
wir-Version								
Eigenwert		1,83	4,20	2,51	1,95	18,20		
% Varianz		2,35	5,39	3,22	2,50	23,37		

Anmerkung: Items mit Faktorladungen kleiner/gleich .30 sind rot gekennzeichnet (KF = Klassenführung, A = Aktivierung, LK = Lernförderliches Klima, M = Motivierung und KV = Klarheit und Verständlichkeit).

## D.3.3 Überblick zu den Fit-Indizes

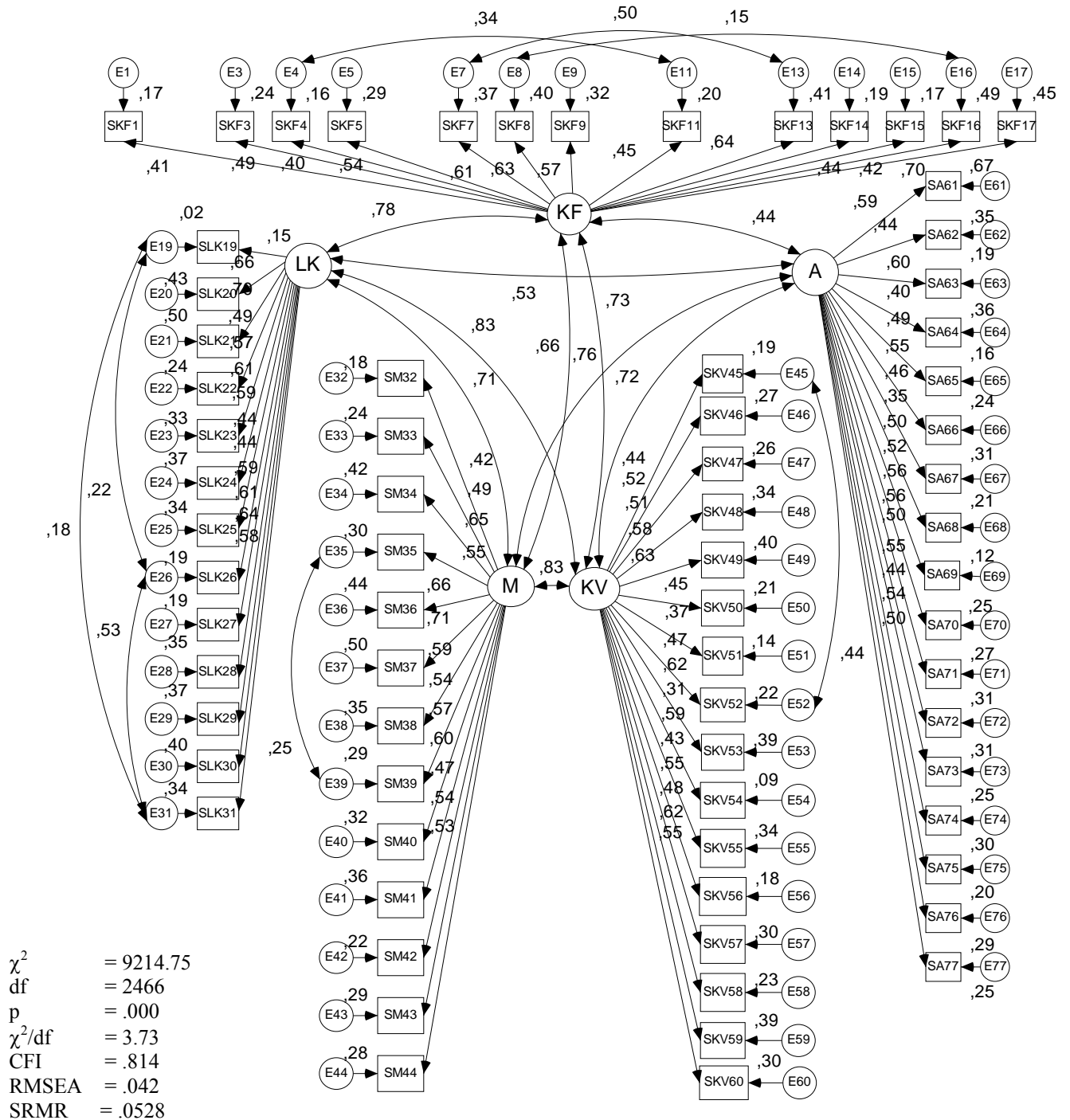
Tabelle D.3-9: Überblick zu den Fitindizes

Fit-Index		Beschreibung	Cut-offs	
$\chi^2$ -Wert	deskriptives Anpassungsmaß			
$\chi^2/df$	deskriptives Anpassungsmaß		$\leq 2.5$	gut
CFI	Comparative Fit Index	<ul style="list-style-type: none"> <li>geringe Sensitivität gegenüber Verteilungsverletzungen und Stichprobengröße</li> </ul>	CFI $\geq .97$	gut
	<i>vergleicht das getestete Modell mit einem restriktiveren Nullmodell bzw. Independence-Modell</i>	<ul style="list-style-type: none"> <li>moderat sensitiv gegenüber einfacher, sehr sensitiv gegenüber komplexer Modellfehlspezifikationen</li> </ul>	CFI $\geq .95$	akzeptabel
RMSEA	Root-Mean-Square-Error of Approximation	<ul style="list-style-type: none"> <li>strenge Prüfgröße</li> <li>sehr sensitiv gegenüber fehlspezifizierten Ladungen</li> </ul>	RMSEA $\leq 0.05$	gut
	<i>betrachtet die Abweichung der beobachteten von der implizierten Varianz-Kovarianzmatrix</i>	<ul style="list-style-type: none"> <li>moderat sensitiv gegenüber einfachen, sehr sensitiv gegenüber komplexen Modellfehlspezifikationen</li> </ul>	RMSEA $\leq 0.08$	akzeptabel
SRMR	Standardized-Root-Mean-Residual	<ul style="list-style-type: none"> <li>sensitiv gegenüber einfacher, moderat sensitiv gegenüber komplexer Modellfehlspezifikation</li> </ul>	SRMR $\leq 0.05$	gut
	<i>kennzeichnet als absoluter Fit-Index die mittlere Abweichung der Residualkorrelationsmatrix</i>	<ul style="list-style-type: none"> <li>geringe Sensitivität gegenüber Stichprobengröße</li> </ul>	SRMR $\leq 0.08$	akzeptabel
Quellen: Backhaus et al. (2006), Bühner (2006), Beauducel & Wittmann (2005) und Schermelleh-Engel, Moosbrugger & Müller (2003)				



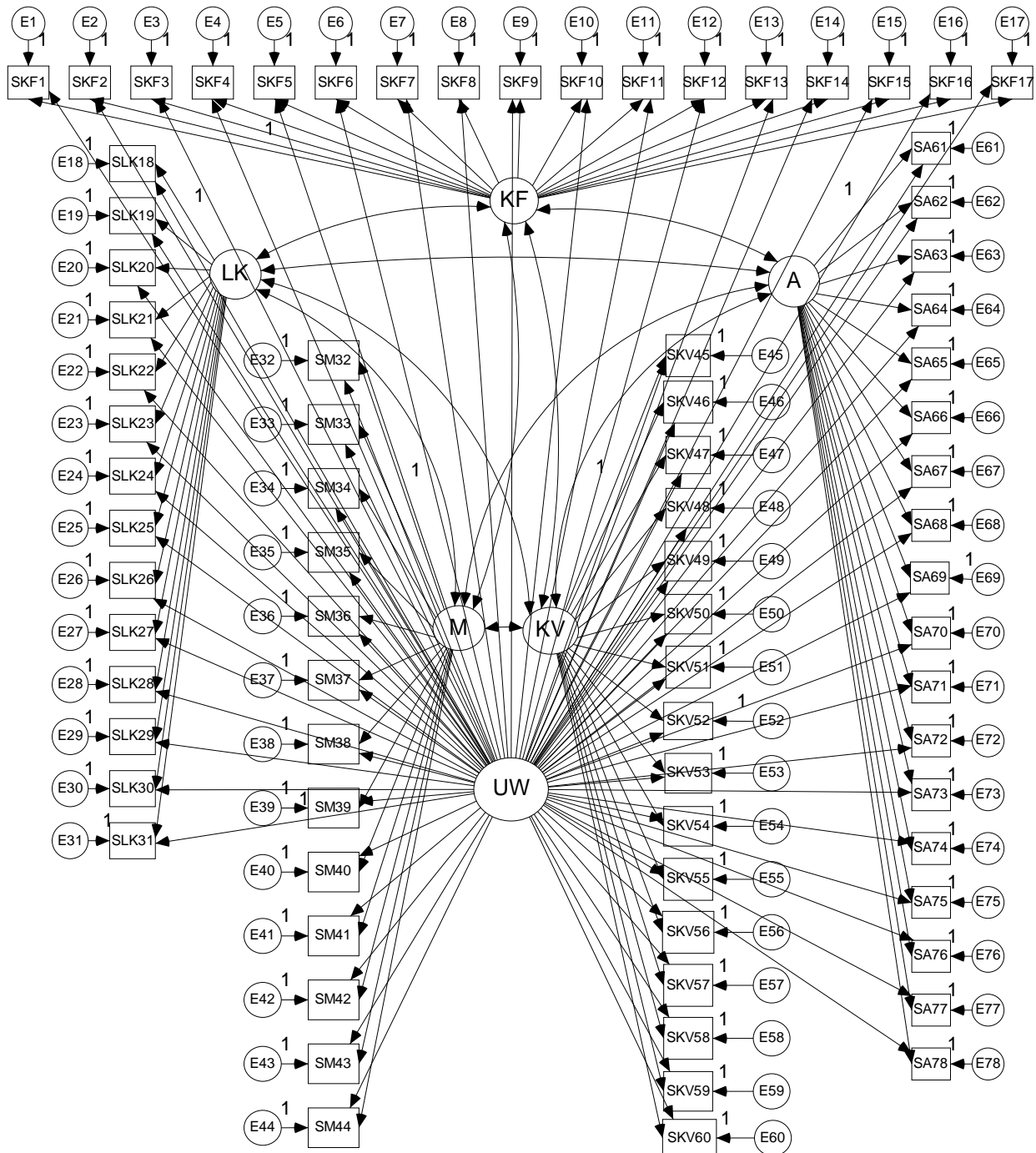
## D.3.4 5-Faktoren-Lösung ohne Items mit einer exploratorischen Faktorladung kleiner .30

Abbildung D.3-2: Konfirmatorische 5-Faktorenlösung ohne Items mit exploratorischer Faktorladung kleiner .30



## D.3.5 Mess- bzw. Strukturmodell der Schmidt-Leiman-Lösung

Abbildung D.3-3: Mess- bzw. Strukturmodell der Schmidt-Leiman-Lösung



## D.4 Zusätzliche Analyse

## D.4.1 ANOVA zur Schüler-Lehrer-Übereinstimmung

Tabelle D.4-10: *1-F-ANOVA zur Abweichung der Schüler- und Lehrerangaben in Abhängigkeit vom Unterrichtsmerkmal*

AV: S\_L\_ABW (Betragsmäßige Abweichung der Schüler-Lehrerangaben hinsichtlich des Itemmittelwerts)

	Quadrat- summe	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	.207	4	.052	2.153	.083
Innerhalb der Gruppen	1.754	73	.024		
Gesamt	1.961	77			

Anmerkung: KF = Klassenführung, LK = Lernförderliches Klima, M = Motivierung, KV = Klarheit und Verständlichkeit, A = Aktivierung

## Anhang E Endgültiger Fragebogen

### E.1 Endgültiger Fragebogen (Version mit männlicher Lehrkraftbezeichnung)

#### Schülerfragebogen zur Unterrichtsrückmeldung

Liebe Schülerin, lieber Schüler!

Deine Meinung zur Unterrichtsstunde ist gefragt!

Die Befragung ist freiwillig und anonym, d. h. notiere bitte  
keine Angaben zu deiner Person auf diesem Fragebogen!

#### Beim Ausfüllen des Fragebogens achte bitte auf Folgendes:

- Du sollst die Unterrichtsstunde anhand einer Reihe von Aussagen einschätzen. Lies dazu jede Aussage aufmerksam durch und wähle die Antwort aus, die Deiner Meinung nach am besten passt. Es geht um Deine eigene Meinung, deshalb gibt es keine richtigen oder falschen Antworten.
- Kreuze das Kästchen an, das zu Deiner gewählten Antwort gehört.
- Kreuze bitte immer nur ein Kästchen in jeder Spalte an!
- Verwende bitte einen dunklen Stift zum Ausfüllen des Fragebogens!

Dieses **Beispiel** zeigt Dir, wie es geht:

#### Gib an, wie sehr Du den folgenden Aussagen zustimmst:

	stimme nicht zu	stimme eher nicht zu	stimme eher zu	stimme zu
Wir arbeiten oft in kleinen Gruppen .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Was wir im Unterricht durchführen, verstehe ich meistens.....	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Wenn Du Deine Antwort ändern möchtest, dann machst Du das angekreuzte Kästchen ungültig, indem Du es ganz ausfüllst und stattdessen das Kästchen mit Deiner neuen Antwort ankreuzt – also so:

Was wir im Unterricht durchführen, verstehe ich meistens.....	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
---	--------------------------	--------------------------	-------------------------------------	-------------------------------------

**Bitte gib nun Deine Meinung zur Unterrichtsstunde an:**

	stimme nicht zu	stimme eher nicht zu	stimme eher zu	stimme zu
<b>Klassenführung</b>				
1 Die Stunde hat pünktlich begonnen.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2 Niemand hat den Unterricht gestört.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3 Ich war die ganze Stunde über aufmerksam und konzentriert.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4 In der Stunde war es so ruhig, dass ich gut arbeiten konnte .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5 Der Lehrer erklärte Aufgaben und Arbeitsblätter so, dass ich gleich wusste, was zu tun ist.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6 Es wurde keine Zeit für Dinge verwendet, die mit dem Lernstoff nichts zu tun haben .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7 Das Klassenzimmer war aufgeräumt.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8 Die Klassenregeln waren mir bekannt.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9 Der Lehrer hat immer alles mitbekommen, was in der Klasse passiert ist.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10 Der Lehrer achtete darauf, dass es gar nicht erst zu Störungen kam....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Lernförderliches Klima</b>				
11 Der Lehrer hat die Klasse zu Beginn der Stunde freundlich begrüßt.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12 Ich konnte entspannt lernen .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13 Mit Fehlern ist der Lehrer so umgegangen, dass ich daraus etwas lernen konnte .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14 Der Lehrer hat sich freundlich zu mir verhalten.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15 Der Lehrer hat die Schüler/innen immer ausreden lassen, wenn sie dran waren .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16 Der Lehrer hat Vorschläge und Ideen der Schüler/innen ernst genommen.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
17 Wenn der Lehrer eine Frage gestellt hat, hatte ich immer ausreichend Zeit zum Nachdenken.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
18 Der Lehrer hat mich Fehler selbst korrigieren lassen.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
19 Nach Fehlern habe ich nicht nur erfahren, <u>dass</u> etwas falsch war, sondern auch <u>warum</u> .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
20 Bei unvollständigen Antworten hat mir der Lehrer Gelegenheit gegeben, die Antworten zu verbessern .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
21 Der Lehrer hat freundlich reagiert, auch wenn ich etwas Falsches gesagt habe .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
22 Der Lehrer hat die Atmosphäre durch humorvolle Bemerkungen aufgelockert .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	stimme nicht zu	stimme eher nicht zu	stimme eher zu	stimme zu
<b>Motivierung</b>				
23 Der Lehrer hat Verbindungen zu anderen Fächern hergestellt.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
24 Der Lehrer hat Beispiele aus dem Alltagsleben gegeben .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
25 Die Aufgaben fand ich abwechslungsreich.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
26 Der Lehrer hat mich mit seinen Beispielen neugierig gemacht, mehr vom Stoff zu erfahren .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
27 Die Aufgaben hatten etwas mit meinen Interessen zu tun.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
28 Der Lehrer hat erklärt, dass der Unterrichtsstoff für das Lernen in anderen Fächern wichtig ist.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
29 Der Lehrer hat gezeigt, dass ich den Unterrichtsstoff auch im Alltag gebrauchen kann .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
30 Der Lehrer hat mir klar gemacht, dass es sich lohnt, mich für dieses Fach anzustrengen.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
31 Bei guten Antworten sind die Schüler/innen von dem Lehrer gelobt worden .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
32 Ich war während des Unterrichts aktiv bei der Sache .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Klarheit und Verständlichkeit</b>				
33 Zu Unterrichtsbeginn hat der Lehrer darauf hingewiesen, was das Ziel dieser Stunde ist.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
34 Der Lehrer hat so laut und deutlich gesprochen, dass ich alles sehr gut verstehen konnte .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
35 Der Lehrer hat mich an Unterrichtsstoff erinnert, den ich früher einmal gelernt habe .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
36 Der Lehrer hat mir Hinweise zu besonders wichtigen Punkten gegeben .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
37 Der Lehrer hat sich so klar und deutlich ausgedrückt, dass ich den Stoff verstanden habe.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
38 Zu Beginn der Stunde hat der Lehrer Stoff aus vergangenen Stunden kurz wiederholt.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
39 Es gab anschauliche Beispiele, die mir das Verstehen des Stoffes erleichtert haben .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
40 Der Lehrer hat am Ende der Stunde die wichtigsten Ergebnisse zusammengefasst.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
41 Der Lehrer hat darauf geachtet, dass ich mich klar und deutlich ausdrücke .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
42 Der Lehrer hat nachgeprüft, ob ich seine Erklärungen wirklich verstanden habe .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
43 Mir war klar, was ich in dieser Stunde lernen sollte .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	stimme nicht zu	stimme eher nicht zu	stimme eher zu	stimme zu
<b>Aktivierung</b>				
44 Der Lehrer hat mir Hinweise gegeben, wie ich besser lernen kann (z.B. zur Arbeitsplatzgestaltung) .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
45 Der Lehrer hat mich dazu aufgefordert, Fragen zu stellen, wenn mir etwas nicht klar ist .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
46 Der Lehrer hat mir Tipps gegeben, wie ich Aufgaben selbstständig lösen kann.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
47 Ich habe meine Arbeitsergebnisse selbst kontrolliert und korrigiert .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
48 Es gab Aufgaben, die ich auf unterschiedliche Weise lösen konnte .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
49 Der Lehrer hat mir Hinweise zur Zeitplanung und Zeiteinteilung beim Lernen gegeben.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
50 Ich habe in der Stunde meine eigenen Arbeitsergebnisse vorgestellt ...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
51 Der Lehrer hat Aufgaben gegeben, die ich selbstständig lösen sollte ...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
52 Ich habe etwas an der Tafel oder am Overhead-Projektor dargestellt...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
53 Ich durfte andere Schüler/innen aufrufen .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
54 Ich habe von mir aus Vorschläge (z. B. zum Inhalt oder Ablauf der Stunde) gemacht .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
55 Ich habe anderen Schülern/innen Fragen zum Unterrichtsstoff gestellt	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
56 Ich habe Lösungswege gesucht und sie der Klasse vorgestellt.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
57 Der Lernstoff wurde durch Abbildungen veranschaulicht (z.B. Mindmap).....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
58 Der Lehrer hat Fragen gestellt, die mich zum Nachdenken angeregt haben.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
59 Der Lehrer hat mich selbst Aufgaben erfinden lassen .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
60 Ich bin während der Unterrichtsstunde mindestens einmal drangekommen.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## **Mein Dank gilt**

...*Dr. Birgit Pikowsky* für ihre kompetente Unterstützung und Betreuung, hilfreichen Rat und wertvolles Feedback in der gesamten Diplomarbeitszeit!

...*Prof. Dr. Manfred Hofer* für die Vermittlung dieser Diplomarbeit, seine fachliche Unterstützung und seinen wohlwollenden Rat!

...*Prof. Dr. Andreas Helmke* für die Möglichkeit der Projektmitarbeit sowie seine Unterstützung während der Fragebogenerstellung und -erhebung!

...*Dr. Wolfgang Wagner* für seine statistischen Tipps und die angenehme Projektzusammenarbeit!

...*Dr. Marten Clausen* für seine inhaltliche und mentale Unterstützung!

...*Dr. Manuel Völkle* für seine methodischen Tipps und *Dr. Caroline Flick* für ihre inhaltlichen Anregungen!

...*Hans-Jürg Liebert* (IFB, Speyer), *Katja Groß-Minor* (IFB, Speyer) sowie *Marianne Hoch* (IFB RFZ, Boppard) für die intensive Unterstützung im Rahmen der Stichprobenakquise!

...*Beate Hofmeister* für ihre fürsorglichen Rat vor und während der Diplomarbeitsphase!

...*Sabrina Schöner*, *Susanne Koudela*, *Ulrike Orth* und *Andreas Vonderschmidt* für ihre hilfreichen Tipps!

...*Katrin Jorke* und *Julia Renz* für ihre tatkräftige Unterstützung bei den Korrekturarbeiten und ihren mentalen Zuspruch!

...meinen *Eltern*, *Geschwistern* und *Hermann* für die Unterstützung während der Diplomarbeitsphase!

...und allen *Lehrkräften und Schülern/innen*, die an den Befragungen teilgenommen und so diese Diplomarbeit erst ermöglicht haben!



## **Eidesstattliche Erklärung**

Hiermit erkläre ich ehrenwörtlich,

1. dass ich die vorliegende Arbeit ohne fremde Hilfe angefertigt habe,
2. dass ich die Übernahme wörtlicher Zitate aus der Literatur sowie die Verwendung der Gedanken anderer Autoren an den entsprechenden Stellen der Arbeit gekennzeichnet habe.

Ich bin mir bewusst, dass eine falsche Erklärung rechtliche Folgen für mich haben kann.

Mannheim, 9. Oktober 2008

---

(Unterschrift)

## **Freiwillige Erklärung**

Ich stimme ausdrücklich zu, dass meine durch Frau Dr. Birgit Pikowsky betreute Diplomarbeit mit dem Titel:

**„Unterrichtsqualität aus der Schülerperspektive:**

**Ein Verfahren zur Unterrichtsentwicklung“**

nach Beendigung der Diplom-Prüfung zu wissenschaftlichen Zwecken im Bereich der Fakultätsbibliothek PE aufgestellt und zugänglich gemacht wird (Veröffentlichungen nach §§ 6 Abs. 1 UrhG), und hieraus im Rahmen des § 51 UrhG zitiert werden kann.

Sämtliche Verwertungsrechte nach § 15 UrhG verbleiben beim Verfasser der Diplomarbeit.

---

(Unterschrift)

Piskol, Kathleen  
(Name, Vorname)

Römerstr. 28a  
69226 Nußloch  
\_\_\_\_\_  
(Anschrift)

Mannheim, 9. Oktober 2008  
(Ort und Datum)